

**Univerzita Karlova**

**Filozofická fakulta**

**Katedra psychologie**



**FILOZOFICKÁ FAKULTA**  
**Univerzita Karlova**

## **Diplomová práce**

Bc. Kateřina Holeyšovská

**Problematika reliability a validity u dynamické diagnostiky**

Reliability and Validity Issues of the Dynamic Assessment

Praha 2017

Vedoucí práce: PhDr. Lenka Krejčová, Ph.D.

## **Poděkování**

Velmi děkuji PhDr. Lence Krejčové, Ph.D. za velikou podporu, vstřícnost a milý přístup při vedení mé diplomové práce i za odborné konzultace, cenné rady a podnětné připomínky. Dále bych chtěla poděkovat Mgr. Ing. Marku Vrankovi za konzultování statistického zpracování dat. Mé poděkování patří také všem odborníkům, kteří s námi ochotně spolupracovali při sběru dat. V neposlední řadě děkuji svým blízkým za trpělivost a podporu.

**Prohlášení:**

*Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.*

*V Praze, dne 11. prosince 2017*

.....

Kateřina Holeyšovská

**Klíčová slova (česky):**

dynamická diagnostika, reliabilita, validita, Dynamická diagnostika kognitivních funkcí u dětí (ACFS), zkušenost zprostředkovaného učení (MLE), hodnotitel

**Keywords (in English):**

dynamic assessment, reliability, validity, Application of Cognitive Functions Scale (ACFS), Mediated Learning Experience (MLE), assessor

**Abstrakt (česky):**

Tato práce se zabývá problematikou reliability a validity metod dynamické diagnostiky. V teoretické části nejprve stručně představujeme specifika dynamické diagnostiky a různé způsoby prověřování reliability a validity v kontextu dynamické diagnostiky. Následně podáváme přehled způsobů prověřování reliability a validity u konkrétních metod dynamické diagnostiky na základě zkušeností převážně ze zahraničních studií. Empirická část si klade za cíl prozkoumat vztah mezi výstupy metody ACFS a charakteristikami hodnotitelů a následně navrhnout optimální přístup k prověřování metod dynamické diagnostiky. Výsledky ukázaly, že hodnotitel byl statisticky významným faktorem pro velikost dosažené změny ve všech subtestech metody ACFS; podrobnější analýza však ukázala, že v každém ze subtestů se významně lišili jiní hodnotitelé. Kvalitativní rozbor nenaznačuje význam konkrétních charakteristik pro rozdíly mezi hodnotiteli.

**Abstract (in English):**

This thesis introduces the issue of reliability and validity of the dynamic assessment methods. In the theoretical part, the author briefly presents specifics of the dynamic assessment and various ways of verifying reliability and validity in the context of dynamic assessment. The text then summarizes ways of verifying reliability and validity for specific methods of the dynamic assessment based on experience from mostly foreign studies. The empirical part of the thesis aims to examine the relationship between the outputs of the ACFS method and characteristics of the assessors and subsequently to propose an optimal approach to the examination of the dynamic assessment methods. The results showed that the assessor was a statistically significant factor to the magnitude of the change in all ACFS scales. However, a more detailed analysis showed that the difference between the assessors varied across subtests. The qualitative analysis did not signify any relevance of specific characteristics to the differences between the assessors.

## OBSAH

Úvod.....	9
TEORETICKÁ ČÁST .....	10
1 Specifika dynamické diagnostiky .....	11
1.1 Hlavní principy a odlišnosti od standardizovaných metod .....	11
1.2 Specifika z hlediska možností standardizace .....	13
2 Reliabilita a validita v kontextu dynamické diagnostiky .....	15
2.1 Způsoby odhadů reliability .....	15
2.1.1 <i>Test – re-test</i> .....	15
2.1.2 <i>Ekvivalence</i> .....	16
2.1.3 <i>Split-half</i> .....	16
2.1.4 <i>Vnitřní konzistence</i> .....	17
2.1.5 <i>Shoda posuzovatelů</i> .....	17
2.2 Zdroje důkazů o validitě .....	19
2.2.1 <i>Obsahová validita</i> .....	19
2.2.2 <i>Kriteriální validita</i> .....	20
2.2.3 <i>Inkrementální validita</i> .....	21
2.2.4 <i>Diferenciální validita</i> .....	22
2.2.5 <i>Ekologická validita</i> .....	22
2.2.6 <i>Validita založená na důsledcích testování</i> .....	23
3 Úloha hodnotitelů v dynamické diagnostice.....	25
3.1 Funkce hodnotitele .....	25
3.1.1 <i>Instrukce</i> .....	25
3.1.2 <i>Zprostředkování</i> .....	26
3.1.3 <i>Hodnocení</i> .....	26
3.2 Způsob intervence .....	26
3.2.1 <i>Strategie zprostředkování</i> .....	27
3.2.2 <i>Zpětná vazba</i> .....	28
3.3 Význam dovedností a charakteristik hodnotitele .....	28

3.3.1	<i>Dovednosti</i> .....	29
3.3.2	<i>Charakteristiky</i> .....	30
4	Metody dynamické diagnostiky a jejich lokalizace .....	31
4.1	Dynamické testování kognitivních funkcí (ACFS) – Lidz .....	31
4.2	Dynamické vyšetření učebního potenciálu (LPAD) – Feuerstein .....	34
4.3	Tzurielova baterie testů .....	35
4.4	Ostatní metody dynamické diagnostiky .....	36
4.4.1	<i>Budoffův přístup</i> .....	36
4.4.2	<i>Metoda stupňovaných pobídek – Campione a Brown</i> .....	37
4.4.3	<i>Guthkeho Lerntest</i> .....	38
4.4.4	<i>Swansonův test kognitivních procesů (S-CPT)</i> .....	38
	EMPIRICKÁ ČÁST .....	40
5	Výzkumný projekt a jeho cíle .....	41
6	Výzkumné otázky .....	42
7	Metody .....	43
7.1	ACFS.....	43
7.2	Dotazník.....	44
8	Výzkumný vzorek.....	45
8.1	Administrátoři metody .....	45
8.2	Vzorek diagnostikovaných dětí.....	46
9	Průběh výzkumu .....	47
10	Výsledky .....	48
11	Diskuse.....	56
	Závěr .....	62
	Seznam použité literatury:.....	63
	Seznam příloh: .....	67

**Seznam použitých zkratk:**

ACFS	Application of Cognitive Functions Scale
CANAL-FT	Cognitive Ability for Novelty in Acquisition of Language as applied to Foreign Language Test
CATM	Children's Analogical Thinking Modifiability
CBDA	Curriculum-based Dynamic Assessment
CCPAM	Children's Conceptual and Perceptual Analogical Modifiability
CITM	Children's Inferential Thinking Modifiability
CMB	Cognitive Modifiability Battery
CSTM	Children Seriation Thinking Modifiability
EOWPVT-R	Expressive One-Word Picture Vocabulary Test-Revised
FTCM	Frame Test of Cognitive Modifiability
LPA	Learning Potential Assessment
LPAD	Learning Potential Assessment Device
MI	modifiability index
MLE	Mediated Learning Experience
RTI	Response to Intervention
S-CPT	Swanson Cognitive Processing Test
SLPT	Series Learning Potential Test



## **Předmluva**

Tato práce navazuje svým tématem na lokalizaci metody ACFS v českém prostředí z roku 2014. Původní název byl „Prověření reliability metody ACFS“ a záměrem výzkumného projektu bylo prověřit reliabilitu metody ACFS v českém prostředí. Diplomový projekt však prošel několika zásadními úpravami. Po zmapování možností sběru dat jsme narazily na skutečnost, že na základě změny v legislativě, která vešla v platnost 1. září 2016, jsou školská poradenská zařízení v současné době natolik vytížená administrativní agendou, že metodu ACFS z důvodu časové náročnosti momentálně prakticky vůbec neadministrují. Z toho důvodu je nereálné nasbírat množství dat potřebné pro původní záměr diplomového projektu. Na základě této skutečnosti jsme byly nuceny přepracovat design diplomové práce. Ta se nebude věnovat prověřování reliability jako takové, nýbrž obecně zmapování různých možností zjišťování reliability a validity u dynamické diagnostiky a navržení optimálního přístupu ke standardizaci metod tohoto typu.

## Úvod

Dynamická diagnostika učebního potenciálu, inspirována především Vygotského teorií zóny proximálního vývoje a Feuersteinovou teorií kognitivní modifikovatelnosti, je slibnou cestou k individuálnímu přístupu ve vzdělávání. V reakci na nedostatky statického testování, dynamická diagnostika nabízí náhled na proces učení a individuální reakce každého testovaného jedince. Administrátor metod dynamické diagnostiky provází testovaného učební situací a navazuje na reakce dítěte přiměřenou intervencí.

Tento způsob práce s sebou z principu přináší řadu otázek. V konfrontaci se statickým způsobem testování jsou to především otázky týkající se metodologie a psychometrie, jako je otázka po reliabilitě a validitě tohoto způsobu testování. Problematika reliability a validity metod dynamické diagnostiky je vzhledem k jejím specifickým velmi důležitým a mezi světovými odborníky často diskutovaným tématem. Přes veškeré diskuse stále existuje mnoho nevyřešených otázek a téma by zasloužilo více studií, které by se jím zabývaly.

V této práci se budeme zmíněné problematice věnovat z několika úhlů pohledu. Nejprve popíšeme specifika dynamické diagnostiky, která by měla být při prověřování psychometrických charakteristik dynamické diagnostiky brána v úvahu, poté se budeme zabývat jednotlivými možnostmi prověřování těchto charakteristik v kontextu dynamické diagnostiky, následně se budeme věnovat tomu, jakou má v tomto procesu úlohu administrátor neboli hodnotitel, a na závěr teoretické části se pokusíme na základě dosavadních studií podat přehled způsobů, kterými jsou metody dynamické diagnostiky ve světě prověřovány. V empirické části se pokusíme lépe prozkoumat úlohu administrátora dynamické diagnostiky pomocí analýzy dat z české lokalizace metody ACFS a charakteristik administrátorů, kteří se na této lokalizaci podíleli. Na závěr na základě dosavadních poznatků zkusíme popsat optimální způsob prověřování dynamické diagnostiky.

## **TEORETICKÁ ČÁST**

# 1 Specifika dynamické diagnostiky

V této kapitole se budeme zabývat specifiky dynamické diagnostiky z hlediska možností standardizace. Nejprve se pokusíme popsat hlavní principy a s nimi spojené odlišnosti, kterými se dynamická diagnostika liší od tradičních standardizovaných metod, a následně vystihnout specifika v kontextu možností standardizace a jejích principů.

## 1.1 Hlavní principy a odlišnosti od standardizovaných metod

Jedním z hlavních specifíků dynamické diagnostiky je její záměr. Jak uvádí Lidz (2003), dynamická diagnostika není alternativou ke standardizovaným metodám. Má zcela odlišné cíle a poskytuje jiné informace. Na rozdíl od standardizovaných metod, které jsou zaměřeny především na odhalení odchylek od normy, je dynamická diagnostika založena na zahrnutí intervence do diagnostického procesu a klade větší důraz na procesy učení než na výsledky. Jejím záměrem je tedy jednak napomáhat změnám v procesu učení s cílem odstraňovat nedostatky, ale také prozkoumávat možnosti a podmínky lepšího výkonu dítěte (Lidz, 2003).

Vedle zahrnutí intervence do diagnostického procesu udává Lidz (2014) jako další hlavní princip dynamické diagnostiky zaměření na proces učení, a to jak na průběh intervence, tak na reakci hodnoceného na tuto situaci a následnou možnost vytváření hypotéz o potenciálně úspěšných intervencích pro daného jedince. K těmto kritériím přidává další, která dohromady označuje jako *kritéria pro dobrou dynamickou diagnostiku*. Zde uvádí následující body:

- „*diagnostický proces zahrnuje relevantní intervenci podle záměru a obsahu hodnocení;*
- *diagnostický proces poskytuje informace o povaze a úrovni reakce testovaného na intervenci;*
- *diagnostický proces umožňuje činit závěry týkající se učebních procesů, které mohou být v průběhu hodnocení měněny;*
- *diagnostický proces poskytuje příležitost k určení míry shody mezi pozorovateli ohledně závěrů a výsledků hodnocení žáka;*

- *diagnostický proces poskytuje informace, které vedou k pozitivním výsledkům pro žáka ve vztahu k budoucímu plánování programu či instrukcím*“ (Lidz, 2014, str. 295).

Tzuriel (2001) podává přehled hlavních rozdílů mezi dynamickou diagnostikou a standardizovanými testy, které srovnává v několika dimenzích. Jmenovitě se jedná o dimenze cíle testování, orientace, kontext testování, interpretace výsledků a povaha úkolů.

**Cílem** dynamického testování je hodnocení změny, zprostředkování, deficitu kognitivních funkcí a hodnocení jiných než rozumových faktorů. Naproti tomu, cílem standardizovaných testů je hodnotit aktuální výkon, porovnávat s vrstevníky a předvídat budoucí úspěch. Výrazné rozdíly jsou také v **zaměření** testování a samotné interpretaci výsledků. Dynamické testování se zaměřuje na proces učení, metakognitivní procesy a porozumění chybám, což je **interpretováno** kvalitativně a v kontextu zprostředkovaného učení. Standardizované testování se orientuje na výsledek a objektivní skóry, které jsou posuzovány objektivně z hlediska průměrných výsledků (Tzuriel, 2001).

Dynamická diagnostika se od statického testování liší také v obecném pohledu na schopnosti jedince. Vnímá schopnosti jako stále se rozvíjející, definuje je jako „*probíhající proces osvojování a upevňování dovedností potřebných pro vysokou úroveň odbornosti*“ (Sternberg & Grigorenko, 2002, str. 3). Statické testování klade důraz na výsledek již rozvinutých dovedností, naproti tomu dynamická diagnostika se zaměřuje na kvantifikaci psychologických procesů zahrnutých v učení a změně. Z dynamického pohledu na schopnosti tak výstupy statických testů kognitivních dovedností, které měří již vyvinuté schopnosti, musejí být neúplné (Sternberg & Grigorenko, 2002).

Další významnou odlišností vyplývající z výše uvedených principů je, v porovnání se standardizovanými metodami, odlišný **výstup** dynamické diagnostiky. Kromě diagnostické klasifikace a dosažené úrovně výkonu zahrnuje nástin překážek pro optimální úroveň výkonu, popis specifických potřeb žáka, doporučení k instrukcím a intervencím a podklady k další péči. Nejdůležitějším výstupem dynamické diagnostiky je žákova reakce na zahrnutou intervenci, která poskytuje informace o takzvaném potenciálu (Lidz, 2014).

Jak podotýká Tzuriel (2000), odlišná forma výstupů či výsledků dynamického testování, především u častého modelu dynamické diagnostiky pre-test – intervence – post-test, může vést k otázce, který z výsledků více reflektuje schopnosti dítěte. Zda pre-test, či post-test. K tomu Tzuriel (2000) uvádí, že na základě teorie, výzkumu i jeho zkušenosti

jsou výkony dítěte v post-tesu přesnějším ukazatelem rozumových schopností dítěte (Tzuriel, 2000).

Dynamická diagnostika nekončí určením skóru post-testu či změny. Proces určování potenciálu žáka Lidz (2014) nazývá screeningem, jelikož se jedná o proces popisný, nikoliv hodnotící. Upozorňuje, že screening není finální fází, ale je důležitý pro rozhodování vzhledem k potřebě dalšího hodnocení. Unikátnost dynamické diagnostiky spočívá v možnosti poskytnutí podkladů, doporučení a informací užitečných pro reálnou intervenci a pro ty, kteří se žákem interagují v běžném prostředí (Lidz, 2014).

Další podstatnou odlišností dynamické diagnostiky od statické, která s výše zmíněnými úzce souvisí, je **kvalita vztahu mezi testovatelem a testovaným**. Na rozdíl od jednostranného vztahu a snahy být co nejvíce neutrální a nezúčastněný v průběhu statického testování, se v průběhu dynamického testování vytváří vztah interaktivní a oboustranný, v případě testování jednotlivce také individualizovaný (Sternberg & Grigorenko, 2002).

Obecně se autoři (např. Lidz, 2014; Tzuriel, 2001) shodují, že hlavním jmenovatelem dynamické diagnostiky je změna. Hodnotitel se snaží intervencí podnítit u testovaného změnu, jejíž povaha a úroveň je následně hodnocena (Lidz, 2014). Průběh dynamické diagnostiky lépe než statické metody koresponduje s procesy učení v reálných životních situacích, a proto mohou kritéria změny v testové situaci sloužit jako ukazatele budoucího kognitivního výkonu (Tzuriel, 2001).

## **1.2 Specifika z hlediska možností standardizace**

Z hlediska možností standardizace bych zde vyzdvihla dimenzi, kterou Tzuriel (2001) pojmenoval jako *kontext testování*. Zatímco kontext statických testů lze charakterizovat jako standardizovaný, strukturovaný a formální, dynamické testování je z principu dynamické a interaktivní a zahrnuje poskytování podpory, pomoci a zpětné vazby.

Jak bylo zmíněno, v dynamické diagnostice je kladen důraz na změnu, což neodpovídá principu standardizace, která je založena neměnných, uniformních podmínkách, jak ji popisuje např. Ferjenčík (2010). Jak navíc uvádí Lidz (2014), způsob, jakým změnu zaznamenáváme a měříme, je stále těžký úkol, ačkoliv existují nějaké možnosti. Dle Ferjenčíka (2010) je standardnost předpokladem k porovnávání dvou výsledků (mezi jedinci, nebo v čase), což bývá u statických psychologických testů často

záměrem. Výsledky jedince jsou zde srovnávány s populačním průměrem. To však není záměrem dynamické diagnostiky. Na rozdíl od statické diagnostiky jsou v rámci dynamické diagnostiky výkony testovaného jedince naopak podněcovány individuální intervencí ke změně (Tzuriel, 2001). To je z principu nutným zásahem do standardních podmínek, a proto není standardizace dynamických metod v tomto pojetí možná.

Standardizace však může být pojata různými způsoby. V užším slova smyslu jako standardnost podmínek testování, nebo jako nastavení hranic pro porovnávání výkonu s výkonem v populaci, či v nejširším slova smyslu jako naplnění standardů testování, což zahrnuje i zajištění vysoké validity a reliability (Urbánek, Denglerová, & Širůček, 2011).

Tzuriel (2000) uvedl, že validita a reliability dynamické diagnostiky jsou dvě hlavní nedořešená témata dynamické diagnostiky, která by měla být prozkoumána. Ačkoliv od té doby bylo publikováno několik studií s touto tematikou, jde o témata stále velmi aktuální.

## 2 Reliabilita a validita v kontextu dynamické diagnostiky

Problematika zjišťování těchto psychometrických hodnot u dynamické diagnostiky je poměrně komplikovaná, jak může být zřejmé ze specifik této metody. Psychometrické údaje tradičních diagnostických metod a jejich hodnota jsou založeny převážně na neměnnosti a stabilitě výstupů hodnocení v čase a možnosti předvídat budoucí výsledky, zatímco dynamická diagnostika vyzdvihuje porovnání změny v průběhu vývoje (Lidz & Peña, 1996).

Myšlenka měření a kvantifikování učeného potenciálu namísto konkrétních naučených dovedností je zajímavá a lákavá, ale měřit spolehlivě a přesně něco, co se neustále rozvíjí a mění, se zdá být těžko realizovatelné (Sternberg & Grigorenko, 2002).

Carver (1974) rozlišuje psychometrickou a edumetrickou dimenzi testů a na základě toho i takzvanou „*psychometrickou reliabilitu*“ a „*edumetrickou reliabilitu*“ (Carver, 1974, str. 514). Termín edumetrická reliabilita vysvětluje jako konzistenci zisku či růstu u jedince, nikoliv jako konzistenci v rozlišování mezi jedinci. Test je považován za psychometricky reliabilní, pokud konstantně rozlišuje mezi dvěma jedinci ve dvou situacích, edumetricky reliabilní, pokud konstantně rozlišuje změnu u jednoho jedince ve dvou situacích (Carver, 1974).

### 2.1 Způsoby odhadů reliability

V této kapitole se pokusíme podat přehled základních způsobů odhadů reliability, a jejich možnosti, limity a přínosy pro dynamickou diagnostiku.

#### 2.1.1 Test – re-test

Jedním z častých způsobů odhadů reliability je metoda test – re-test. Reliabilita je zde pojata jako stabilita v čase a je založena na předpokladu neměnnosti měřených charakteristik testované osoby. Tento předpoklad však neodpovídá zcela realitě, jelikož vlastnosti lidské psychiky mohou v souvislosti s vnitřními či vnějšími podmínkami kolísat či se měnit (Urbánek et al., 2011).

Navíc, jak bylo zmíněno v předchozí kapitole, intervence a následná změna je v dynamické diagnostice základním a žádoucím principem. V dynamické diagnostice nemůžeme, a ani nechceme, předvídat výstupy post-testu na základě pre-testu, tak jak je to



žádoucí u tradičního testování (Lidz & Peña, 1996). Ze stejných důvodů nemůžeme ani považovat post-test za nástroj odhadu reliability dynamické diagnostiky (Camilleri & Botting, 2013).

Tento způsob odhadu reliability může být použit pro statický mód dynamických testů neboli pro odhad reliability pre-testu, jak je tomu například u LPAD (Learning Potential Assessment Device) či jiných dynamických metod. V tomto kontextu může dosahovat vysokých hodnot, jelikož testové baterie se v mnoha případech skládají z upravených verzí statických testů. V kontextu dynamické diagnostiky je však důležité především hodnocení reliability změny či přínosu intervence, nikoliv samotného výstupu testu (Sternberg & Grigorenko, 2002).

Byly navrženy i způsoby odhadů reliability, které zahrnují větší počet alternativních pre-testů a post-testů, které by umožňovali hodnotit změnu opakovaně (Carver, 1974). Konstantní úroveň rozdílu mezi pre-testem a post-testem napříč alternativními formami by se dala interpretovat jako vysoká reliability změny (Carver, 1974).

### 2.1.2 Ekvivalence

Ekvivalence je způsob odhadu reliability na základě porovnání paralelních forem testu administrovaných těsně po sobě. Je založen na předpokladu, že jakýkoli rozdíl ve výstupech obou forem testu představuje chybu měření. Koeficient reliability zde odpovídá korelaci výsledků paralelních forem testu (Urbánek et al., 2011).

Tento způsob odhadu reliability je u dynamické diagnostiky, s ohledem na charakter dynamických metod, možný ve formě ekvivalentních pre-testů i post-testů a následného porovnání změny a zlepšení v čase, jako je tomu např. studie autorů Hasson, Dodd, a Botting (2012), ve které byly pro účely hodnocení reliability vytvořeny čtyři paralelní verze dané metody.

### 2.1.3 Split-half

Metoda „split-half“ funguje na podobném principu jako ekvivalence, s tím rozdílem, že místo ekvivalentní formy využívá rozdělení testu na dvě poloviny. Výhodou je, že není třeba ekvivalentní forma, která u řady testů neexistuje a navíc je obtížné ji vytvořit, nevýhodou je naopak potřeba velkého počtu testových položek (Urbánek et al., 2011).

Z toho vyplývá, že v případě, kdy konkrétní test dynamické diagnostiky disponuje dostatečným počtem položek, můžeme tuto metodu odhadu reliability použít, pokud rozdělíme zvlášť položky pre-testu a post-testu a následně vyhodnotíme konstantnost změny obdobně jako u předchozích způsobů.

#### 2.1.4 Vnitřní konzistence

Mezi nejpoužívanější odhady reliability obecně patří v současnosti Cronbachův koeficient alfa určující vnitřní konzistenci testu na základě korelace dvojic položek (Urbánek et al., 2011).

Tento způsob výpočtu koeficientu reliability je možné použít i u metod dynamické diagnostiky, jak ukazují například studie autorů Brandão, Cadime, Ribeiro, a Viana (2014), Camilleri a Botting, (2013) nebo De Beer (2010). Swanson (1995) doporučuje vnitřní konzistenci jako vhodný nástroj u metod, kde v případě re-testu očekáváme změnu, jako je tomu u dynamické diagnostiky.

#### 2.1.5 Shoda posuzovatelů

Shoda posuzovatelů se používá jako ukazatel reliability v situacích, kdy hodnocení výsledků či výstupů testové situace probíhá se zapojením interpreta (Urbánek et al., 2011). Proto je nejčastěji používanou metodou odhadu reliability dynamické diagnostiky právě shoda posuzovatelů. Byla použita v několika studiích, které reliability dynamické diagnostiky prověřovaly, například studie autorů Binger, Kent-Walsh, a King (2017), Kelley (2017), Hansen, Steultjens, a Satink (2009), Camilleri a Botting (2013).

Ve studii zkoumající dynamickou diagnostiku metodou stupňovaných pobídek u dětí s poruchou řeči (Binger et al., 2017) byla shoda posuzovatelů zjišťována za použití videonahrávek z testování. Vyškolení výzkumníci hodnotili výstupy dynamické diagnostiky skrze videonahrávky a následně byla počítána shoda mezi jejich hodnoceními pomocí Cohenova koeficientu kappa (Binger et al., 2017).

Kelley (2017) ve své studii počítal koeficient reliability jako procentuální shodu přidělených skóre dvou nezávislých hodnotitelů u každé položky. Posuzovatelé hodnotili položky rovněž z videozáznamu procesu testování a na základě detailní příručky pro skórování s uvedenými příklady (Kelley, 2017).

Hansen et al. (2009) pro odhad reliability vyhodnocovali shodu devíti vyškolených hodnotitelů. V tomto případě byly porovnávány výstupy jednoho hodnotitele s dalšími osmi hodnotiteli, kteří vyhodnocovali výstupy za použití videonahrávek.

Stejně tomu bylo i ve studii autorů Camilleri a Botting (2013), kde kvalifikovaní odborníci pro odhad reliability vyhodnocovali dynamickou diagnostiku prostřednictvím videonahrávek ze sezení na základě sepsaných instrukcí a vodítek pro konkrétní metodu.

Na základě informací o principech dynamické diagnostiky a výše zmíněných studií se ukazuje být shoda posuzovatelů pro odhad reliability dynamické diagnostiky přínosnou metodou. Nahrávání diagnostických sezení na video umožňuje nezávislé hodnocení dvěma či více posuzovateli a následné srovnání skóru. To je z hlediska individuálního posuzování deficitů či změny klíčové.

Otázka způsobu zjišťování reliability je velmi aktuální i u nového konceptu „Response to Intervention“ (RTI), který reprezentuje stejný konstrukt jako dynamická diagnostika. Oba koncepty, RTI i dynamická diagnostika, jsou založeny na změně. Jednak na změně u dítěte a také na změně přístupu k dětem (Grigorenko, 2009).

Jak uvádí Dykeman (2006), systém RTI reprezentující hodnotící metody a strategie založené na reakci dítěte na intervenci by měl být hodnocen z hlediska reliability zvlášť v jednotlivých fázích procesu. Bylo by vhodné tedy věnovat pozornost nejen reliability přidělování skóru, ale také reliability identifikace problému, zvolení vhodné intervence, procesu intervence a doporučení (Braden, Kubiszyn & Ortiz, 2006, podle Dykeman, 2006).

Odhady reliability dynamické diagnostiky často ukazují uspokojivé výsledky, pokud jsou administrovány ve statickém modu. Kritickým tématem zůstává reliability shody posuzovatelů ve vyvozených důsledcích a volbě potřebné intervence, které jsou stěžejními tématy dynamické diagnostiky (Haywood & Wingenfeld, 1992).

Je zřejmé, že proces intervence hraje v dynamické diagnostice důležitou roli. Tento proces se může lišit mimo jiné zvoleným typem zprostředkování, jehož reliability lze rovněž posuzovat pomocí shody posuzovatelů (Tzuriel & Samuels, 2000 podle Tzuriel, 2000). Rolí hodnotitele v procesu intervence u dynamické diagnostiky se budeme podrobněji zabývat v kapitole 3 „Úloha hodnotitelů v dynamické diagnostice“.

## 2.2 Zdroje důkazů o validitě

Validita neboli „míra, ve které empirické důkazy a teorie podporují interpretaci testových skóre při doporučeném způsobu použití testu“ (*Standardy pro pedagogické a psychologické testování*, 2001, str. 17) je nejdůležitějším hlediskem pro hodnocení testu. Nehodnotíme validitu testu jako takového, ale validitu interpretace skóre. Způsoby, jakým budou skóre využity, vytvářejí specifický rámec a ovlivňují interpretaci skóre. Při validizaci testu je tedy třeba hledat takové typy důkazů, které podporují interpretaci pro specifický účel testu (*Standardy pro pedagogické a psychologické testování*, 2001).

Z tohoto důvodu je důležité při rozhodování, které typy důkazů o validitě daného testu jsou vhodné, brát v potaz hlavní principy a účely dynamické diagnostiky, kterým byla věnována kapitola 1.

### 2.2.1 Obsahová validita

Obsahová validita, přesněji řečeno důkazy validity založené na obsahu testu, dokládá vztah obsahu testu a konstruktů, který má test měřit (*Standardy pro pedagogické a psychologické testování*, 2001). Může prokázat obsahové pokrytí určité domény, to však není pro prokázání skutečnosti, že metoda měří nějaký atribut dostačující. Proto je třeba pro poskytnutí důkazů o validitě metody srovnání s kritériem měřeného atributu, které je nezávislé na daném testu (Urbánek et al., 2011). Tomu se více věnujeme v části „Kriteriální validita“.

Přesto můžeme tento způsob důkazu validity najít v některých studiích prověřujících validitu metod dynamické diagnostiky (např. Baek & Kim, 2003; Hansen et al., 2009), často společně s dalšími druhy důkazů. Obsahová validita může být zaměřena na několik oblastí, které můžeme rozdělit na domény adekvátnost konstruktů a interpretací skóre, vhodnost a reprezentativnost škál, dostatečné informace pro administraci, použitelnost v klinické praxi a kvalita obsahu (Hansen et al., 2009).

Validita zaměřená na adekvátnost konstruktů je také nazývána **konstruktová validita**. Rozumíme tím hodnocení, zda měřený atribut odpovídá konstruktům v rámci relevantní psychologické teorie (Urbánek et al., 2011). Například korelace skóre testu „Cognitive Processing Test“ (S-CPT) s běžným experimentálním měřením pracovní paměti, nebo zkoumání konstruktů modelu S-CPT pomocí faktorové analýzy (Swanson, 1995).

Autoři Baek a Kim (2003) se ve své studii vedle obsahové validity, kterou dokládají rozbořením podoblastí použitého testu dynamické diagnostiky, opírají také o zjevnou validitu.

**Zjevná validita**, neboli „face validita“, je podle Urbánka et al. (2011) míra, do jaké je možné probandem odhadnout účel testování, jinými slovy to, co má metoda měřit. Obecně není tato metoda dokazování validity příliš uznávaná a také není příliš často používána.

Baek a Kim (2003) dokládají zjevnou validitu, společně s obsahovou validitou, a to rozhovory s profesorem, studenty oboru zaměřeného na hodnocení vzdělání a učiteli mateřské školy, kteří metodu posuzovali.

Zjišťování obsahové validity metod dynamické diagnostiky může být přínosné z hlediska zdůraznění komplexnosti takové péče a případné potřeby upřesnění definic či operacionalizace postupů pro vedení vhodné intervence (Hansen et al., 2009).

### 2.2.2 Kriteriační validita

Kriteriační validita, ve standardech z roku 2001 pod názvem „*důkazy založené na vztahu k jiným proměnným*“ (*Standardy pro pedagogické a psychologické testování*, 2001, str. 21), je chápána jako zdroj potvrzení validity založený na vztahu skóru testu s nějakým vnějším kritériem (*Standardy pro pedagogické a psychologické testování*, 2001).

Nalezení vhodného a reliabilního kritéria je pro mnoho psychologických atributů obtížné a v některých případech ani neexistuje (Urbánek et al., 2011). Validita dynamické diagnostiky může být navíc demonstrována pouze částečnou korelací s vnějším kritériem, jelikož dosažení kritéria často není cílem dynamické diagnostiky (Hasson et al., 2012).

Do této kategorie spadá i tak zvaná **prediktivní validita**, která vyjadřuje možnost výsledků testu predikovat hodnoty jiného kritéria naměřené v budoucnu (*Standardy pro pedagogické a psychologické testování*, 2001).

Poměrně často se jako kritérium pro prediktivní validitu používá index založený na školní klasifikaci coby kritérium školního úspěchu. Nicméně toho kritérium je samo o sobě problematické z hlediska objektivity, reliability i validity, a proto je často do značné míry nekvalitní (Urbánek et al., 2011). Školní úspěch je navíc spíše produkt než proces, což je argumentem pro nevhodnost tohoto kritéria pro validizaci dynamické diagnostiky, která klade velký důraz na proces (Sternberg & Grigorenko, 2002).

Kritérium prediktivní validity je důležitým a často používaným zdrojem důkazů validity testů dynamické i statické diagnostiky, jak dokládá například přehledová studie, která ukázala, že dynamická diagnostika předpovídá budoucí úspěch s podobnou přesností jako statické testování, a navíc také poukazuje na její přidanou hodnotu viz „Inkrementální validita“ (Caffrey, Fuchs, & Fuchs, 2008).

Důležitými otázkami pro zjišťování prediktivní validity jsou: jaké hodnoty chceme predikovat, které dovednosti jsou pro tyto hodnoty reprezentativní, můžeme výstup referenčního testu zobecnit na hodnotu kritéria, nemáme k dispozici citlivější zdroj informací pro stanovení kritéria, než je referenční test (Caffrey et al., 2008)?

Potvrzování validity pomocí vnějšího kritéria může mít také řadu omezení. Například omezení limitovaného výběru nástrojů měřících požadované kritérium, případně zjednodušené měření kritéria a nedostatečnost komplexního pojetí, či nedostatečné zaznamenávání hodnot kritéria v přirozeném prostředí školní třídy (Swanson, 1995).

U kritéria prediktivní validity u dynamické diagnostiky musíme brát také v potaz, že predikce může být úspěšná pouze do té míry, do které jsou realizována doporučení k intervenci vycházející z dynamické diagnostiky (Hasson et al., 2012).

### 2.2.3 Inkrementální validita

Inkrementální neboli přírůstková validita metody spočívá v dokázání přínosu oproti jiným zdrojům. Může být pojata jako přínos dané metody v případě konkrétní testové situace, nebo obecně ve vztahu k určitému kritériu vzhledem k odlišnému testu (Urbánek et al., 2011).

U dynamické diagnostiky můžeme uvažovat o přínosu oproti statickému testování například v tom, že vysvětluje a předpovídá významně větší procento budoucích úspěchů jak ve výkonových, tak kognitivních testech. Může předpovídat unikátní výkony, které nejsou zachyceny a předpovídaný tradičním testováním (Caffrey et al., 2008). Jak uvádí Haywood a Wingenfeld (1992), na základě dynamické diagnostiky můžeme získat informace, které nejsou dostupné ze standardizovaných metod, a to především díky konceptu podněcování změny a následného hodnocení efektu.

Budoucí výzkum dynamické diagnostiky by se proto neměl zaměřovat pouze na srovnávání se statickými testy, ale měl by spíše lépe prozkoumat a poskytovat informace o tom, co může dynamická diagnostika přinést nad rámec tradičního testování (Caffrey et al., 2008).

#### 2.2.4 Diferenciální validita

Diferenciální (kontrastní) validita je založena na opačném principu než korelace, tedy na odlišnosti atributu měřeného daným testem od jiného atributu (Urbánek et al., 2011).

U metody dynamické diagnostiky CANAL-FT (Cognitive Ability for Novelty in Acquisition of Language as applied to Foreign Language Test) bylo pro hodnocení diferenciální validity použito srovnání s měřením inteligence, které prokázalo, že se metoda liší od měření krystalické i fluidní inteligence (Sternberg & Grigorenko, 2002).

V případě dynamické diagnostiky měřící schopnost učit se slova to může být rozlišení mezi poruchou řeči a odlišností jazyka, navíc v porovnání se statickým hodnocením. Výsledky studie ukázaly, že dynamická diagnostika, konkrétně kombinace EOWPVT-R („Expressive One-Word Picture Vocabulary Test-Revised“) a MI („modifiability index“), rozlišuje mezi dětmi s typickým vývojem a dětmi se sníženými řečovými schopnostmi lépe než metody statického testování, které často falešně vyhodnocují poruchy řeči. Výsledky této studie mají význam pro klinické využití metod dynamické diagnostiky při hodnocení řečových schopností dětí z různých jazykově odlišných prostředí. Studie také zdůrazňuje limity tohoto způsobu prověřování validity, které spočívají ve velké závislosti na testované populaci (Peña, Iglesias, & Lidz, 2001).

#### 2.2.5 Ekologická validita

Kritéria ekologické validity, jako například pozorování ve školní třídě, prohlášení rodiče či zpráva učitele, mohou být užitečná pro hodnocení validity metody dynamické diagnostiky z hlediska jejího přínosu a využitelnosti výstupů v praxi a porozumění výsledkům v širším kontextu (Peña et al., 2001).

Proto se zdá být vhodná kombinace ekologické validity prověřující metodu v přirozeném kontextu a dalších způsobů důkazů validity, jako je tomu například ve studii autorů Camilleri a Botting (2013). Autoři této studie odvozují ekologickou validitu od řešení úkolů daného testu v přirozeném kontextu a současně předkládají důkazy souběžné a prediktivní kritériální validity na základě silné korelace s vnějšími kritérii (Camilleri & Botting, 2013).

Sternberg a Grigorenko (2002) pro ekologickou validitu dynamického testování argumentují tím, že pokud je součástí testování pedagogická intervence, pak samotné testování má didaktický význam.

### 2.2.6 Validita založená na důsledcích testování

Důsledkům testování v souvislosti s důkazy validity byla věnována pozornost až v nedávné době. Cílem takové validizace je určení pravděpodobnosti, že použití testu přinese užitek, který byl při zadávání testu očekáván, tedy prokázat tvrzení a shromáždit důkazy o prospěšnosti testu. Například pokud má test sloužit k rozřazení osob, měla by validizace prokázat prospěšnost následného rozřazení (*Standardy pro pedagogické a psychologické testování*, 2001).

Vztáhneme-li to na dynamickou diagnostiku, měla by tedy validizace založená na důsledcích testování pravděpodobně prokázat, že diagnostický proces včetně intervence a navrhovaných postupů byl pro daného jedince prospěšný.

Hodnocení validity psychodiagnostických metod obecně vyžaduje ověřování pro různé účely testování a skupiny probandů, navíc test není validní pouze z jednoho hlediska, ale je třeba ověřovat mnoho různých aspektů validity (Urbánek et al., 2011). U dynamické diagnostiky je třeba validizovat jednotlivé cíle či účely zvlášť a pomocí jiného kritéria, z čehož vyplývá, že validizace metod dynamické diagnostiky je složitějším a komplexnějším procesem než u statického testování (Tzuriel, 2000).

Studie prokazující validitu dynamické diagnostiky se navíc většinou zaměřují na validitu v krátkodobém horizontu, málokteré prokazují validitu interpretace interaktivního hodnocení z hlediska dlouhodobého (Haywood & Tzuriel, 1992 podle Haywood & Wingenfeld, 1992).

Dle autorů Caffrey, Fuchs, a Fuchs (2008) existují na prověřování reliability a validity metod dynamické diagnostiky mezi odborníky dva odlišné pohledy, které se odvíjejí od rozdílných přístupů k dynamické diagnostice. Těmito přístupy jsou „**klinicky orientovaná** dynamická diagnostika“, v jejímž průběhu examinátor často mění přístup a způsob zpětné vazby za účelem poskytnutí vhodného způsobu intervence a maximalizování studentova úspěchu, a „**výzkumně orientovaná** dynamická diagnostika“, která typicky zahrnuje standardizované testy a samostatnou fázi, kde vede examinátor lekci učení. Povaha poskytované zpětné vazby v těchto odlišných přístupech se odráží v pohledu na standardizaci. Mnoho zastánců klinicky orientované dynamické diagnostiky záměrně odmítá standardizaci procesu, jelikož zastávají názor, že standardizace jde proti teoretickým myšlenkám a principům dynamické diagnostiky a znemožnila by poskytnout skutečně individualizovanou intervenci. Zastánci výzkumně orientované dynamické



diagnostiky naopak věří, že standardizace a technická adekvátnost jsou nezbytné pro vytvoření hodnotného nástroje (Caffrey et al., 2008).

Vedle těchto přístupů také existuje možnost administrovat nestandardizované metody dynamické diagnostiky výzkumným způsobem, tedy s oddělenými fázemi pre-test – intervence – post-test, kdy jsou pro pre-test a post-test použity paralelní nebo stejné verze testu. Tak je tomu například i v metodě ACFS, kde je administrace jednotlivých subtestů rozdělena na fáze pre-test – intervence – post-test. Pro pre-test a post-test je použita stejná verze testu, ale intervence využívá odlišné nástroje i postupy. Pro odlišné fáze jsou zvlášť podrobně popsány postupy administrace i intervence. Průběh intervenční fáze je u této metody do značné míry závislý na interakci s dítětem, nicméně manuál poskytuje vodítka ve formě kontrolního seznamu pro zprostředkování a formulář v záznamovém archu pro zaznamenání hodnotitelových postupů intervence. To umožňuje snadnější výzkumné využití (Lidz & Krejčová, 2014).

S podobným přístupem se setkáme i u metod D. Tzuriela, které lze využít jak klinicky se začleněním intervence kdykoliv v průběhu testování podle potřeb klienta, tak i výzkumně s oddělením jednotlivých fází. Na rozdíl od ACFS jsou pro tyto účely u metod vytvořené paralelní verze záznamových archů pro různý způsob diagnostické práce (např. Tzuriel & Klein, 1990; Tzuriel, 1990; Tzuriel & Galinka, 2002).

Haywood a Wingenfeld, (1992), kteří se zabývali interaktivním hodnocením jako výzkumným nástrojem, jsou toho názoru, že čím reliabilnější a validnější bude měření změny, tím lépe a užitečněji budeme moci dynamickou diagnostiku měřící tuto změnu ve výzkumu používat. Důležitým aspektem pro zvažování vhodných psychometrických charakteristik u dynamické diagnostiky je však účel a kontext testování, jak již bylo zmíněno výše. Pokud je naším záměrem využití dynamické diagnostiky pro výzkumné účely, mohou mít větší důležitost jiné charakteristiky než při klinickém použití.

### 3 Úloha hodnotitelů v dynamické diagnostice

V předchozích kapitolách jsme se blíže věnovaly specifikům a možným způsobům prověřování reliability a validity dynamické diagnostiky. Jak bylo zmíněno, kvalita vztahu hodnotitele a testovaného je jedním z důležitých specifík dynamické diagnostiky a má v tomto diagnostickém procesu důležitou funkci. Z toho důvodu je úloha hodnotitele důležitým tématem i v kontextu prověřování reliability a validity metod dynamické diagnostiky.

V této kapitole se proto budeme věnovat nejen náplni a průběhu práce hodnotitele, ale především důležitosti a významu této role v procesu dynamického testování v kontextu zjišťování reliability a validity.

#### 3.1 *Funkce hodnotitele*

Jak je zřejmé již z principů dynamické diagnostiky, od tradičních přístupů se dynamická diagnostika rolí hodnotitele podstatně odlišuje. Na rozdíl od převážně pasivní a neutrální role zapisovatele v tradičním testování, má hodnotitel v dynamickém testování aktivní funkci intervenujícího, který zapojuje dítě do učební interakce a pokouší se podněcovat pozitivní změny (Lidz & Peňa, 1996).

Testující poskytuje testovanému zpětnou vazbu, implicitní či explicitní v závislosti na typu testování, podává instrukce za účelem pomoci testovanému zlepšit výkon, a je schopen hodnotit vývoj procesu řešení problému v kontextu reakce na intervenci (Sternberg & Grigorenko, 2002).

##### 3.1.1 Instrukce

Míra, do jaké se mohou schopnosti zlepšit či rozvinout, nezáleží pouze na latentní kapacitě žáka, ale také na způsobu a druhu poskytnutých instrukcí, které pomáhají tuto latentní kapacitu rozvíjet. Způsob, jakým je test dynamické diagnostiky administrován v souvislosti se způsobem zadávání úkolů pro stanovení kritéria, může být jednou z intervenujících proměnných ovlivňujících korelaci testu s vnějším kritériem. Z toho vyplývá, že způsob, jakým hodnotitel test či úkoly administruje, může mít podstatný význam pro validizaci dynamického testu (Sternberg & Grigorenko, 2002).

### 3.1.2 Zprostředkování

Zprostředkování je nedílnou součástí procesu dynamické diagnostiky. Jak uvádí Tzuriel (2000), hodnotitel by měl v procesu dynamické diagnostiky vynaložit úsilí pro rozvoj dítěte. Z toho je zřejmá důležitost a do určité míry i náročnost procesu intervence, která je jednou z funkcí hodnotitele v dynamické diagnostice.

Například v metodě dynamické diagnostiky LPAD hodnotitel interaguje s dítětem na základě konceptu MLE (Mediated Learning Experience). Podstatou MLE je kvalitní interakce mezi učícím se dítětem a podnětem zprostředkovaná kompetentním dospělým, který podněty pro dítě vybírá, mění, zdůrazňuje a interpretuje. Nedostatečné množství, nebo nesprávný typ zprostředkování učiva či jiných podnětů může mít za následek snížený učební potenciál, nebo klasifikování dětí jako podprůměrných, znevýhodněných či mentálně opožděných (Kozulin & Rand, 2000).

Možnostem a způsobům intervence a zprostředkování se více věnujeme v podkapitole „Způsob intervence“.

### 3.1.3 Hodnocení

Skórování a vyhodnocování dynamického testu je také jednou z proměnných týkajících se korelace výsledku daného testu s vnějším kritériem, v širším pohledu tedy kritériální validity (Sternberg & Grigorenko, 2002).

Záměrem dynamické diagnostiky není samotné měření zisku následujícího po intervenci, ale především vyhodnocení toho, jak a proč které děti reagují na trénink za účelem navržení příhodné intervence (Caffrey et al., 2008; Elliott, 2003). Jak uvádí Sternberg a Grigorenko (2002), hodnotitelovo vnímání testovaného dítěte v průběhu procesu je skutečně významným aspektem dynamické diagnostiky, nutně se tedy dotýká i problematiky reliability a validity.

## 3.2 *Způsob intervence*

Zde se budeme věnovat způsobu, jakým hodnotitel v procesu dynamické diagnostiky provádí intervenční fázi a strategiím, které k tomu používá. Zaměříme se především na srovnání efektu jednotlivých strategií.

### 3.2.1 Strategie zprostředkování

Vědecké studie zabývající se dynamickou diagnostikou dětí se speciálními vzdělávacími potřebami se často zaměřují kromě ukazování efektu zprostředkování na kognitivní výkony dětí právě na identifikaci toho, která technika zprostředkování je úspěšnější v produkování změny (Tzuriel, 2000).

Tzuriel (2000) uvádí, že zkušenosti ukazují rozdíly v chování i v kognitivních dovednostech dětí se stejnými charakteristikami, jako jsou věk, pohlaví a IQ, které tudíž potřebují i odlišné strategie učení. Hodnotitel tedy musí individuálně vyhodnotit vhodnost strategie pro konkrétní dítě.

Různé strategie intervence a jejich efekt zkoumali i Kester, Peña, a Gillam (2001) kteří srovnávali přístupy MLE, přímé instrukce a smíšený přístup. Výzkum ukázal, že všechny strategie zprostředkování byly významně úspěšnější než neposkytnutí žádné intervence. Zároveň strategie MLE se ukázala být přibližně třikrát a smíšený přístup přibližně dvakrát více efektní než přímé instrukce, a to i u dětí z kulturně a jazykově odlišného prostředí. To autoři vysvětlují tím, že přístup MLE, na rozdíl od přímých instrukcí, zahrnuje zprostředkované porozumění pravidlům v širším kontextu a naučení se strategie řešení. Přístup založený na přímých instrukcích se zaměřuje převážně na naučení konkrétním novým dovednostem, zatímco MLE je zaměřené na zprostředkování učení metakognitivních strategií i v kontextuálně odlišných úlohách prostřednictvím kombinace základních komponentů MLE, jimiž jsou záměrnost a reciprocita zprostředkovaného učení, přesah obsahu zprostředkovaného učení v kontextu relevantním pro dané dítě, podpora kompetentnosti ve vytváření plánu směřujícímu k danému záměru a zprostředkování významu učení (Kester et al., 2001).

Otázka, zda by včleněná intervence, neboli učící fáze procesu dynamické diagnostiky měla být více kvalitativní, individualizovaná a zaměřená na dítě, nebo více kvantitativní, standardizovaná a psychometrická, je jednou ze zásadních otázek dynamické diagnostiky týkající se zjišťování reliability a validity (Kozulin & Rand, 2000).

Jak již bylo zmíněno v předchozí kapitole „Reliabilita a validita v kontextu dynamické diagnostiky“, povaha intervence se často odvíjí od přístupu k dynamické diagnostice, nebo od jejího účelu. Při klinickém způsobu administrace metody, kdy je důraz především na individualizovaný přístup v reakci na potřeby konkrétního dítěte, jsou na hodnotitele kladeny vysoké nároky na odbornost a schopnost porozumění, aby mohl adekvátně vyhodnotit obtíže a vhodně přizpůsobit intervenci. Ve výzkumně orientovaném

přístupu je fáze učení striktně oddělena a může být lépe standardizovaná (Caffrey et al., 2008). Jak bylo popsáno výše v této práci, možnosti zjišťování reliability a validity se způsobem intervence úzce souvisejí, je však nutné brát v potaz i účel a kontext testování.

Standardizovaná intervence může být vhodnější pro výzkumné účely, nicméně individualizovaná intervence se ukázala být efektivnější v podněcování změny, zvláště u dětí se speciálními vzdělávacími potřebami (např. Missiuna & Samuels, 1989).

### 3.2.2 Zpětná vazba

Studie zkoumající příčinu velkých interindividuálních rozdílů u dětí ve schopnosti profitovat z tréninku dynamické diagnostiky v úkolu analogického usuzování poukázala souvislost mezi zlepšením dítěte a typem tréninku (Stevenson, Hickendorff, Resing, Heiser, & de Boeck, 2013). Zlepšení v analogickém usuzování v rámci dynamické diagnostiky bylo větší, pokud následovalo po tréninku založeném na stupňování pobídek než po tréninku založeném na zpětné vazbě k výsledku. Metoda stupňování pobídek zahrnuje na rozdíl od samostatné zpětné vazby i instrukce týkající se metakognice a strategie řešení problému (Stevenson et al., 2013). Toto zjištění podporuje i další studie, která prokázala větší účinnost intervence zahrnující zpětnou vazbu ke strategii řešení než intervence, která zahrnovala zpětnou vazbu pouze k výsledku (Luwel, Foustana, Papadatos, & Verschaffel, 2010).

Na základě tohoto zjištění se zdá být způsob, jakým hodnotitel provádí intervenci v rámci dynamické diagnostiky, jedním z důležitých faktorů ovlivňujících rozdíly mezi dětmi nejen v dosažené úrovni změny v rámci vyšetření, ale také změny v přístupu k řešení problémů obecně, tedy i mimo rámec vyšetření (Stevenson et al., 2013).

### 3.3 Význam dovedností a charakteristik hodnotitele

Již bylo zmíněno, že dynamická diagnostika klade na hodnotitele vysoké nároky, a to především z hlediska jeho odbornosti a s tím spojených profesních dovedností. Z toho důvodu se zde věnujeme právě dovednostem a dalším charakteristikám hodnotitele, které jsou v procesu dynamické diagnostiky významné, a které by tudíž mohly potenciálně souviset s možnostmi prověřování reliability a validity metod dynamické diagnostiky.

### 3.3.1 Dovednosti

Jak uvádějí Tzuriel a Haywood (2002), potřebné profesní dovednosti hodnotitele dynamické diagnostiky jsou trénovány prostřednictvím intenzivních workshopů, a přesto hodnotitel musí do značné míry vykonávat subjektivní úsudek v několika oblastech dynamického testování. Hodnotitel musí posoudit, které kognitivní funkce jsou oslabené, jaký způsob zprostředkování je a jaký není vhodný, kdy je a kdy není zprostředkování potřebné a jak správně vyhodnotit a interpretovat změnu mezi pre-testem a post-testem (Tzuriel & Haywood, 2002). Beyer a Magnusson (2003, podle Hansen et al., 2009) uvádějí, že vysoká míra subjektivity implikuje sníženou reliabilitu metody. Z tohoto důvodu má pro prověřování charakteristik dynamické diagnostiky velice důležitý význam shoda mezi hodnotiteli.

Porozumění hodnotitele pro to, co je třeba ve vztahu k dítěti udělat pro dosažení lepšího výkonu, je klíčovou schopností rámující kognitivní výkon dítěte (Sternberg & Grigorenko, 2002). Hodnotitel navíc musí dbát na to, aby dítěti bez záměru neposkytoval neverbální vodítka pro řešení úkolu, aby byl schopen správně vyhodnotit, zda dítě bylo schopné úkol vyřešit samo, případně jaká vodítka byla dítěti nápomocná (Camilleri & Botting, 2013).

Důležitost role hodnotitele v procesu dynamické diagnostiky a jeho postupů, způsobů a intenzity intervence jsou zdůrazněny i u metody ACFS. Hodnotitel by měl jednak mít dovednosti umožňující mu využívat zkušenost zprostředkovaného učení, která je východiskem pro interakci mezi hodnotitelem a testovaným u této metody, a jejímž záměrem je společným učením s dítětem rozvíjet jeho kognitivní funkce a podnítit strukturální kognitivní změny, to znamená dovést dítě k aktivnímu používání postupů přemýšlení a strategií řešení úkolů (Lidz & Krejčová, 2014).

Pro zprostředkování podnětů dítěti je třeba umět použít strategie jako je upoutání pozornosti změnou četnosti, pořadí a intenzitou podnětů, propojení se známým kontextem či významem. Hodnotitel by měl dále používat motivační aspekty, jako je zvědavost, napětí, výzva, a snažit se rozvinout kognitivní funkce dítěte (Tzuriel, 2000).

Vysoké nároky na hodnotitele dynamické diagnostiky spočívají i v nutnosti nepřetržitého reagování na dítě, zvažování jeho možností, potřeb, aktuální motivace, přemýšlení o charakteru úkolů i prostředí. Důležité jsou také schopnosti zjistit na základě sledování rozdílů ve výkonu dítěte mezi pre-testem a post-testem informace o procesu učení, o pokrocích dítěte a proměnách způsobu jeho práce, a v neposlední řadě dovednost

reflexe vlastní práce v průběhu intervence a rozeznání a popsání jakými postupy lze podnítit rozvoj dítěte. Hodnotitel musí v průběhu intervence reflektovat mnoho různých proměnných, které mají v procesu dynamické diagnostiky důležitou funkci (Lidz & Krejčová, 2014).

Užívání metody ACFS je z tohoto důvodu podmíněno zaškolením ve způsobu práce s ACFS a česká verze je doplněna o škálu „Forma a intenzita intervence“, která slouží hodnotiteli k reflexi svého chování a práce v průběhu intervence (Lidz & Krejčová, 2014).

### 3.3.2 Charakteristiky

Charakteristiky hodnotitele byly v souvislosti s reliabilitou metody prověřovány například ve studii autorů Hansen et al. (2009). Autoři prověřovali reliabilitu metody dynamické diagnostiky na základě shody posuzovatelů pomocí výpočtu koeficientu Kappa a následné analýzy možných souvislostí s charakteristikami hodnotitelů, jako jsou uplynulá doba od ukončení studia, nebo zkušenosti klinické praxe. Analýza neprokázala, že by tyto charakteristiky ovlivňovaly výsledky Kappa koeficientu měřícího shodu hodnotitelů. Koeficient Kappa v této studii svědčí o vysoké míře shody mezi hodnotiteli v několika oblastech. Pro hodnocení výkonnostních skóre se koeficienty reliability pohybovaly od dostačující shody (59%) po vysokou míru shody (85%). Pro fázi výstupu dynamické diagnostiky byly koeficienty shody poměrně vysoké v pěti oblastech, kterými byla doporučení vhodné intervence, doporučení vhodných kompenzačních postupů, směřování péče, určení míry instrukcí a určení míry vedení. Slabá míra shody hodnotitelů se ukázala v jedné z měřených oblastí, kterou byla míra poskytované zpětné vazby (Hansen et al., 2009).

Výsledky studie mimo jiné poukazují na důležitost srozumitelnosti definic a operacionalizace nástrojů určených k vedení přiměřených intervencí (Whyte & Hart, 2003). Autoři studie také zdůrazňují důležitost tréninku v metodě a doporučují minimální délku tréninku dva dny, s následným extra tréninkem pro hodnotitele, u kterých se ukáže slabá míra shody v hodnocení s ostatními (Hansen et al., 2009).

Dynamická diagnostika obecně vyžaduje intenzivní trénink administrování metody i interpretování výsledků, jelikož klinická intuice, dovednosti a odbornost hodnotitele mají v tomto přístupu velký význam (Grigorenko, 2009).

## 4 Metody dynamické diagnostiky a jejich lokalizace

V této kapitole se pokusíme podat přehled možností, jakými je prověřována reliabilita a validita metod dynamické diagnostiky. Zaměříme se především na metody používané v České republice a zkušenost s prověřováním těchto metod ze zahraničních studií.

### 4.1 *Dynamické testování kognitivních funkcí (ACFS) – Lidz*

Na základě dat z české lokalizace metody ACFS není možné stanovit reliabilitu pomocí koeficientu vnitřní konzistence, jelikož metoda se skládá z více jednoúkolových škál, ani pomocí metody test – re-test, protože mezi pre-testem a post-testem očekáváme změnu podnícenou intervencí. Validitu je z principu metody obtížné měřit, neboť zde hraje roli velké množství proměnných (Lidz & Krejčová, 2014).

Studie zkoumající validitu metody ACFS provedená v Americe na 26 dětech, z nichž 21 byly děti se speciálními vzdělávacími potřebami, ukázala signifikantní zlepšení mezi pre-testem a post-testem u škál Třídění, Přejímání perspektivy, Verbální plánování a Doplnění sekvenčních vzorců (Shurin, 1999). Ačkoliv interpretace těchto výsledků je omezená, jelikož studie postrádá kontrolní skupinu, jsou tyto výsledky interpretovány jako podpůrné tvrzení pro konstruktovou validitu jednotlivých škál. Obě škály, u kterých se neprokázalo signifikantní zlepšení, se týkaly paměti. Tento náález je možné interpretovat v kontrastu s první studií ACFS autorky C. Lidz, která byla prováděna na populaci nadprůměrně nadaných dětí, kde výsledky ukázaly signifikantní zlepšení i v subtestech týkajících se paměti. Možná interpretace tedy je, že metoda ACFS by na základě výsledků v paměťových subtestech mohla být validní pro rozlišování dětí se speciálními vzdělávacími potřebami a dětí nadprůměrně nadaných (Shurin, 1999).

U subtestů Krátkodobá vizuální paměť, Verbální plánování a Doplnění sekvenčních vzorců, které pro úspěch v post-testu vyžadují plné zaměření pozornosti a koncentrace na úkol, ukázala studie významné spojení skóru post-testu s chováním dítěte v průběhu fáze intervence, což poukazuje na reliabilitu škály Pozorování chování dítěte. Nejsilněji korelovala se skórem v post-testu flexibilita, která je popsána jako míra, do jaké se dítě samo opravuje či zkouší alternativní řešení úkolu, signifikantní korelace se ukázala rovněž u vytrvalosti a frustrační tolerance (Shurin, 1999).



Jiná americká studie zkoumající validitu a reliabilitu škály Pozorování chování dítěte metody ACFS zahrnovala pozorování 25 dětí z mateřské školy pro děti s vývojovými poruchami (Aranov, 1999). Validita byla zkoumána na základě korelace hodnocení výzkumníkem a učitelem či logopedem. Bylo posuzováno chování dítěte ve školní třídě a v průběhu logopedické intervence. Reliabilita byla stanovena pomocí shody posuzovatelů na základě skóre hodnotitele a výzkumníka, který posuzoval nahrané videozáznamy. Ve studii byla reliabilita prověřena také pomocí konzistence chování dětí ve dvou různých situacích, která ukázala signifikantně pozitivní korelaci. Výsledky celkově poukázaly na velmi silnou shodu mezi hodnotiteli především u hodnocení seberegulace, interakce, motivace a odezvy, o něco slabší míru shody u hodnocení vytrvalosti a flexibility (Aranov, 1999).

Další americká studie se zaměřila na prověření subtestů metody ACFS Doplnění sekvenčních vzorců a Krátkodobá vizuální paměť (Malowitzky, 2001). Studie byla provedena na 30 normálně se vyvíjejících předškolních dětech ve věku 3-5 let, které byly rozděleny na kontrolní a experimentální skupinu. Výsledky neukázaly žádný efekt zlepšení u kontrolní skupiny, která neabsolvovala zprostředkované učení. Naopak u experimentální skupiny, která podstoupila zprostředkované učení během intervenční fáze, se ukázalo signifikantní zlepšení v subtestu Doplnění sekvenčních vzorců. V subtestu Krátkodobé vizuální paměti studie neukázala významné zlepšení, což je diskutováno v souvislosti s vysokými skóre v pre-testu a skóre blízkými maximálním hodnotám v post-testu. Výsledky nicméně podporují validitu těchto škál pro testovanou populaci (Malowitzky, 2001).

Validitu tří subtestů ACFS Sekvenční sluchová paměť, Přejímání perspektivy a Verbální plánování prověřovala další americká studie (Bensoussan, 2002). Opět zde byla porovnávána experimentální skupina se zprostředkovanou intervencí a kontrolní skupina bez intervence. U subtestů Sekvenční sluchová paměť a Přejímání perspektivy výsledky ukázaly signifikantní zlepšení u experimentální skupiny a neukázaly signifikantní zlepšení u kontrolní skupiny, což poukazuje na absenci efektu pouze na základě tréninku a podporuje validitu intervenční fáze pomocí zprostředkovaného učení. Subtest Verbální plánování se na základě zlepšení i u kontrolní skupiny ukázal být více náchylný k efektu tréninku (Bensoussan, 2002).

Ve studii prověřující validitu metody ACFS pro populaci nizozemských dětí byly srovnány skóre ACFS se skóre statických testů. Škály Třídění a Doplnění sekvenčních

vzorců byly srovnány s matematickými dovednostmi, paměťové škály byly srovnány s jazykovými dovednostmi skrze pochopení slov, vyprávění příběhu a porozumění příběhu. Škála Sekvenční sluchová paměť signifikantně korelovala se statickým testem vyprávění příběhu, nicméně výsledky se významně nelišili v pre-testu a post-testu. U ostatních škál se neprokázala signifikantní korelace, což je interpretováno tak, že škály ACFS měří jiné dovednosti než statické testy (Van Der Aalsvoort & Lidz, 2007).

Dynamická diagnostika je často také užívána k testování dětí z odlišného kulturního prostředí. Na vhodnost použití metody ACFS pro takto odlišnou populaci poukazuje i studie prokazující nezávislost zlepšení výkonu po intervenci na kulturním prostředí, ze kterého dítě pochází. Efekt zprostředkovaného učení byl prokázán na různorodém vzorku dětí žijících v různých zemích a s odlišným migračním statutem. Autoři proto považují metodu ACFS za vhodnou pro použití ve Španělsku i Německu i u dětí z odlišného jazykového či kulturního prostředí a doporučují studii rozšířit o další země (Wiedl, Mata, Waldorf, & Calero, 2014).

Validita španělské verze ACFS byla prověřena v longitudinální studii, kde bylo 87 předškolních dětí hodnoceno metodou ACFS a po dvou letech byla znovu hodnocena jejich inteligence, metakognice a učební přístup. Na základě metody ACFS byl pro každé dítě vypočítán index učebního potenciálu, který byl následně srovnáván s kritériem získaným testováním ve výše zmíněných oblastech. Výsledky s ohledem na zlepšení v testovaných oblastech ukázaly, že učební potenciál slouží jako důkaz prediktivní validity metody ACFS (Mata & Calero, 2014).

V průběhu lokalizace ACFS pro české prostředí byla provedena studie v rámci diplomového projektu A. Drápalové (2014), která porovnávala výkon experimentální skupiny 25 dětí, u které byla provedena intervence, a srovnávací skupiny 25 dětí srovnatelných z hlediska pohlaví, věku i kognitivních schopností, u které měly děti místo intervence k dispozici podnětový materiál. Tato studie ukázala, že obě skupiny jsou srovnatelné v pre-testu a u experimentální skupiny došlo k významnému zlepšení mezi pre-testem a post-testem ve všech subtestech kromě Verbálního plánování. Tyto výsledky podporují myšlenku, že intervenční fáze testování má pozitivní vliv na kognitivní rozvoj dětí. Ve srovnávací skupině došlo ke zlepšení ve čtyřech subtestech, kterými byli Sekvenční sluchová paměť, Vizualní paměť, Doplnění sekvenčních vzorců a Přejímání perspektivy, což může být vysvětleno přirozenou tendencí učit se, nebo vzhledem k identickému materiálu v pre-testu a post-testu vlivem paměti. Zlepšení experimentální

skupiny mezi pre-testem a post-testem bylo vůči srovnávací skupině významně větší, což je argumentem pro to, že ACFS dokáže měřit potenciál kognitivního rozvoje dětí a je tedy přínosným diagnostickým nástrojem (Drápalová, 2014).

#### **4.2 Dynamické vyšetření učebního potenciálu (LPAD) – Feuerstein**

Tzuriel (2001) zdůrazňuje důležitost reliability metody LPAD a obecně metod dynamické diagnostiky s ohledem na charakter metody a význam role hodnotitele, nicméně podotýká, že studií, které se o prověření reliability LPAD pokusily, je velmi málo.

Vaught a Haywood (1990 podle Tzuriel, 2001) prověřovali reliabilitu dvou testů z LPAD na základě shody hodnotitelů pomocí videonahrávek. Byly hodnoceny deficity kognitivních funkcí, množství a způsob potřebného zprostředkování učení a přenos učebních procesů neboli reakcí na zprostředkování. Celková shoda mezi hodnotiteli byla relativně nízká, stejně tak shoda v hodnocení zprostředkování, nicméně shoda u hodnocení deficitů kognitivních funkcí byla vyšší, okolo 80 %.

Tzuriel a Samuels (2000 podle Tzuriel, 2001) ukázali ve své studii vyšší reliabilitu LPAD na základě shody posuzovatelů u hodnocení deficitu kognitivních funkcí 87-88 %, pro kategorii hodnocení reakce na intervenci a vhodného typu zprostředkování byla shoda nižší, podobně jako u předchozí studie.

Testy Ravenovy progresivní matrice a Test komplexní figury jsou původně statickými metodami, které byly adaptovány na dynamickou verzi jako součást LPAD a jsou tedy standardizovány v jejich statické podobě (Tzuriel, 2001).

V rámci studie zkoumající učební potenciál u nadaných dětí a dětí s průměrnou inteligencí u španělské populace, byly prověřeny tři subtesty LPAD z hlediska reliability a prediktivní validity. Jednalo se o Test prostorového učení, Vzory ze šablon a Organizér. Reliabilita hodnocená na základě vnitřní konzistence testů a test–re-testové reliability ve statické formě dosahovala hodnot 0,7-0,95. Všechny tři testy také ukázaly vysokou míru predikce IQ skóru, naproti tomu IQ skóry nebyly významnými prediktory výstupů dynamické diagnostiky. Tyto výsledky podporují reliabilitu a prediktivní validitu těchto testů pro účely identifikování nadaných dětí (Calero, Belen, & Robles, 2011).

Elliott (2003) uvádí LPAD jako příklad metody klinického přístupu k dynamické diagnostice, která většinou používá formát, kdy nejsou striktně oddělené fáze testování a intervence, ale zprostředkované učení je zahrnuto v průběhu procesu. Ačkoliv intervence vychází z ucelené Feuersteinovy teorie, forma a zaměření intervence jsou ovlivněny

administrátorem a přizpůsobeny individuálním podmínkám dítěte a okolnostem, tudíž se může významně lišit u jednotlivých testujících. Proto tato metoda nepoužívá normy, nesrovnává testované mezi sebou a nesnaží se standardizovat intervenci ani prověřovat reliabilitu a validitu na základě tradiční vědecké metodologie (Elliott, 2003).

#### **4.3 Tzurielova baterie testů**

Tzuriel (2001) uvádí validitu a reliabilitu jednotlivých metod, které do jeho baterie testů pro mladší děti patří, a odkazuje na řadu dalších studií. Reliabilita je většinou vyjádřena pomocí Cronbachova koeficientu alfa a důkazy validity jsou založené na empirických datech.

U testu „Children's Analogical Thinking Modifiability“ (CATM) byla reliabilita počítána pomocí Cronbachova koeficientu alfa, který měl pro pre-test hodnotu 0,72 a pro post-test hodnotu 0,90 (Tzuriel & Klein, 1985). Validita CATM jako metody měřící kognitivní modifikovatelnost byla prověřena na testování čtyř skupin dětí. U znevýhodněných a běžných dětí se ukázal dle očekávání větší zisk z intervence než u dětí se speciálními vzdělávacími potřebami a dětí s mentální retardací (Tzuriel & Klein, 1985).

Test „Children's Inferential Thinking Modifiability (CITM) byl prověřen v několika studiích. Reliabilita CITM vypočítaná pomocí Cronbachova koeficientu alfa má ve fázích pre-testu, post-testu i u fáze zprostředkování hodnotu 0,82-0,90 (Tzuriel, 1992 podle Tzuriel, 2000). Validita této metody byla prověřena například z hlediska efektivity zprostředkování pro znevýhodněné děti, kdy výsledky studie ukázaly vyšší míru zlepšení znevýhodněných dětí než u nadaných, zvláště ve více komplexních úkolech (Tzuriel, 1989 podle Tzuriel, 2000).

U dalších testů baterie pro mladší děti Tzuriel (2001) také uvádí hodnoty koeficientu reliability Cronbachovo alfa a odkazuje na studie podporující validitu těchto metod. Hodnoty Cronbachova koeficientu pro fázi pre-testu a post-testu jednotlivých testů nabývají hodnot 0,85 a 0,90 pro test „Frame Test of Cognitive Modifiability“ (FTCM), 0,87, 0,85 a 0,70 pro „Children Seriation Thinking Modifiability“ (CSTM), pro „Cognitive Modifiability Battery“ (CMB) pro subtest řazení 0,62 a 0,64, pro subtest reprodukce vzorců 0,91 a 0,94, pro subtest analogie 0,84 a 0,77, pro subtest sekvence-I 0,88 a 0,90 a pro subtest sekvence-II 0,69 a 0,67 (Tzuriel, 2001).

Koeficient reliability u metody „Children's Conceptual and Perceptual Analogical Modifiability“ (CCPAM) byl počítán zvlášť pro dvě verze. U verze „Closed Analogies“

nabyl u subtestu pojmové analogie hodnot 0.73 pro pre-test a 0.84 pro post-test, u subtestu percepční analogie 0,82 pro pre-test a 0,89 pro post-test, pro verzi „Construction Analogies“ měl hodnoty pro pojmové i percepční analogie 0,83 a 0,84 (Tzuriel & George, 2009).

#### **4.4 Ostatní metody dynamické diagnostiky**

Níže uvádíme stručný přehled metod dynamické diagnostiky a způsoby zjišťování jejich validity a reliability, které jsou známé ve světě, ale v České republice nejsou běžně užívány. Jedná se o Budoffův přístup, Metodu stupňovaných pobídek, Lerntest a Swansonův test.

Lidz (1991) uvádí, že pro Budoffův přístup a přístup autorů Campione a Brown je charakteristické standardizování intervenční fáze a model test – intervence – re-test. Metody dynamické diagnostiky těchto autorů jsou tak více kontrolovatelné a je snadnější zjišťovat jejich psychometrické vlastnosti (Lidz, 1991).

##### **4.4.1 Budoffův přístup**

Budoffův přístup nazvaný „Learning Potential Assessment“ (LPA) zahrnuje více známých standardizovaných metod upravených do dynamické podoby. Jedná se například o „Kohs Learning Potential Task“ „Raven Learning Potential Test“ „Picture World Game“ a „Series Learning Potential Test“ (SLPT) (Lidz, 1991).

Jejich reliabilita je hodnocena jako dostačující na základě vnitřní konzistence či způsobu test – re-test. Hodnoty vnitřní konzistence jsou vysoké a korespondují s hodnotami verzí statických testů, způsob zjišťování reliability test – re-test byl použit pro pre-testy a pohybuje se v hodnotách 0,51-0,95. Neexistuje však žádné hodnocení reliability měřené změny (Sternberg & Grigorenko, 2002).

Sternberg a Grigorenko (2002) podotýkají, že důkazů pro validitu Budoffova přístupu je velmi málo a studie, které se zabývaly prověřením prediktivní validity toho přístupu, nejsou průkazné, ačkoliv jsou slibné. Prediktivní validita byla prověřována například hodnocením školního úspěchu ve srovnání s hodnocením učitele, nebo statickými IQ testy (Sternberg & Grigorenko, 2002).

Jiná studie zaměřená pouze na jeden test z Budoffova přístupu ukázala, že prediktivní validita SLPT vůči kritériu školního úspěchu se ukázala být vyšší pro post-test než pro pre-test u skupiny dětí s nižším IQ, na rozdíl od skupiny dětí s vyšším IQ. Navíc

nebyla nižší než u testu obecných inteligenčních schopností. Výsledky však nebyly statisticky signifikantní (Babad & Budoff, 1974).

#### 4.4.2 Metoda stupňovaných pobídek – Campione a Brown

Sternberg a Grigorenko (2002) uvádějí, že Metoda stupňovaných pobídek je díky standardizované intervenci novým přístupem k dynamické diagnostice. Vzhledem ke standardizovanému testování bez nutnosti speciálního tréninku testujících a jednoduché administraci je metoda vhodná a často používaná pro výzkumné účely. Prediktivní validita této metody byla prověřena například ve studii, která prokázala, že rozdíl mezi pre-testem a post-testem v této metodě je lepším prediktorem úspěchu než statické testování (Campione 1989 podle Sternberg & Grigorenko, 2002).

Metoda stupňovaných pobídek byla také prověřena ve studii srovnávající stupňované pobídky a MLE. Studie prokázala navzdory kvalitativním odlišnostem efektivitu obou přístupů pro zvyšování porozumění ve čtení (Naeini, 2014).

Sternberg a Grigorenko (2002) uvádějí, že navzdory standardizovanému přístupu, psychometrické vlastnosti této metody, jako jsou test–re-testová reliabilita a vnitřní konzistence, nebyly dostatečně prozkoumány. Uvádějí pouze jednu studii, která hodnotila reliabilitu změny, a nikoliv reliabilitu úkolů. Tato studie ukázala, že množství instrukcí potřebných pro splnění úkolu bylo adekvátní u dvou rozdílných ale podobných úkolů, což svědčí o reliabilitě měření učebního potenciálu u různých úkolů založených na stejném principu (Ferrara et al. 1986 podle Sternberg & Grigorenko, 2002).

Novější studie prověřovala reliabilitu metody dynamické diagnostiky založené na stupňovaných pobídkách experimentem, při kterém výzkumníci vyškolení a trénovaní v administraci i intervenční fázi testu vedli sezení, které bylo nahráváno na videokameru. Sezení trvalo 60 minut, probíhalo v uzavřené výzkumné místnosti a výzkumníci byli požádáni o dodržení standardních procedur testování. Prostřednictvím videonahrávek byla následně sezení skórována jinými výzkumníky a byla posuzována shoda hodnotitelů pomocí výpočtu Cohenova koeficientu kappa. Hodnoty koeficientu kappa 1,0 pro fázi testování a 0,95 pro fázi pobídek svědčí o velmi vysoké reliabilitě hodnocení měření (Binger et al., 2017).

#### 4.4.3 Guthkeho Lerntest

Sternberg a Grigorenko (2002) uvádějí, že vzhledem k tomu, že se jedná o test vytvořený v Německu, je většina studií psána v cizím jazyce a tedy nedostupná. Z dostupných studií zmiňují výzkum podporující validitu testu na základě většího zlepšení ve výkonu u experimentální skupiny a prediktivní validitu post-skórů ve srovnání s kritériem školních známek a hodnocení učitelem (Guthke & Wingenfeld, 1992 podle Sternberg & Grigorenko, 2002), a také studii ukazující, že Lerntest je v predikci osvojení znalostí i aplikaci znalostí v komplexních úkolech a v predikci výkonu lepší než statické testy, v predikci výkonu v matematice a jazykových úkolech srovnatelný se statickým testováním (Guthke et al. 1997 podle Sternberg & Grigorenko, 2002).

Důkazy pro validitu Lerntestu byly prezentovány také výzkumem, který ukázal, že Lerntest je efektivním indikátorem intelektuálních schopností testovaného (Guthke, Beckmann, & Dobat, 1997 podle Neubauer, Grabner, Freudenthaler, Beckmann, & Guthke, 2004).

Co se týká reliability této metody, nenašly jsme žádné dostupné údaje.

#### 4.4.4 Swansonův test kognitivních procesů (S-CPT)

Metoda S-CPT byla vyvinuta pro zjišťování, zda se u dětí se speciálními vzdělávacími potřebami obecně vyskytuje deficit v pracovní paměti, zda se děti s deficitem v pracovní paměti liší od ostatních dětí s poruchami učení, a do jaké míry je modifikovatelná pracovní paměť (Sternberg & Grigorenko, 2002).

Validita této metody byla prověřována několika způsoby v rámci standardizace na 1611 probandech v deseti amerických státech a dvou provinciích Kanady. Metodu administrovali školní psychologové a absolventi psychologie, kteří prošli tréninkovým kurzem v této metodě. Konstruktová validita byla podpořena silnou korelací skóre jednotlivých subtestů SCP-T s pracovní pamětí. Analýza prokázala signifikantní efekt zlepšení výkonu mezi pre-testem a post-testem S-CPT, což je také důkazem pro konstruktovou validitu testu. Korelace S-CPT s výkonovými testy poukázala ve srovnávání s krátkodobou pamětí na silnou divergentní validitu. Prediktivní validita testu byla podpořena zlepšením predikce schopnosti čtení na základě skóre dynamické diagnostiky (Swanson, 1995).

Reliabilita SCP-T byla odhadována na základě vnitřní konzistence. Cronbachův koeficient alfa byl počítán zvlášť pro fáze vstupního testování, změny, udržení a pro

jednotlivé subtesty, pohyboval se pro všechny fáze v hodnotách 0,72 – 0,92 a pro subtesty 0,82 – 0,95, což svědčí o vysoké reliabilitě. Celkově je odhad reliability vyšší pro test jako celek, než pro jednotlivé subtesty (Swanson, 1995).



## **EMPIRICKÁ ČÁST**

## 5 Výzkumný projekt a jeho cíle

Jak bylo zmíněno v předmluvě, výzkumný projekt této práce prošel podstatnou změnou. Původním záměrem bylo, v návaznosti na lokalizaci metody ACFS pro Českou republiku z roku 2014, prověřit reliabilitu této metody. Vzhledem k okolnostem a současným možnostem však není tento záměr realizovatelný. Níže se proto budeme věnovat pouze novému projektu této diplomové práce.

Cílem tohoto diplomového projektu je zmapovat souvislosti výstupů dynamické diagnostiky s jednotlivými hodnotiteli, kteří metodu administrovali, a navrhnout optimální model pro ověření reliability a validity metod dynamické diagnostiky. Pokusíme se tak učinit na základě analýzy dat z lokalizace metody ACFS v českém prostředí a vybraných charakteristik hodnotitelů, kteří se na této lokalizaci podíleli. Zaměříme se na normalitu rozložení výsledků v testované populaci a rozložení u hodnotitelů a prozkoumáme vztah mezi hodnotiteli a skóry výsledné změny u jednotlivých škál metody ACFS. Pokusíme se najít souvislosti mezi charakteristikami jednotlivých hodnotitelů a těmito výsledky.

Naším záměrem je lépe prozkoumat a popsat specifika použité metody v českém prostředí a na základě toho se pokusit navrhnout nejvhodnější možný způsob prověření metod dynamické diagnostiky.

## 6 Výzkumné otázky

Výzkumné otázky tohoto projektu formulované na základě charakteristik metody ACFS a obecných principů dynamické diagnostiky jsou následující:

1. Existuje vztah mezi konkrétními hodnotiteli a skóry výsledné změny u jednotlivých škál metody ACFS?
2. Objevují se u hodnotitelů rozdílné tendence napříč jednotlivými škálami metody ACFS?
3. Které charakteristiky se častěji ukazují u hodnotitelů, u kterých se objevily výrazné tendence v hodnocení?

Předpokládáme normalitu rozložení výsledků ve skóru změny v testované populaci a rovnoměrné rozložení výsledků ve skóru změny u jednotlivých hodnotitelů. Celkový předpoklad je, že napříč hodnotiteli není významný rozdíl v hodnocených skórech a charakteristiky hodnotitelů nemají pro hodnocení dynamické diagnostiky velkou roli.

Hlavní hypotéza, kterou budeme statisticky testovat, je formulována následovně:

**H<sub>0</sub>:** Mezi průměry skóre změny dílčích subtestů není u probandů testovaných různými administrátory významný rozdíl.

**H<sub>A</sub>:** Průměrné skóry změny dílčích subtestů se významně liší u probandů testovaných různými administrátory.

## 7 Metody

V této kapitole blíže představíme českou adaptaci ACFS, jelikož pro účely projektu byla použita data z lokalizace této metody. Stručně popíšeme jednotlivé škály, fázi procesu intervence a způsob vyhodnocení.

Dále popíšeme metodu dotazníku, pomocí které jsme získaly údaje o charakteristikách jednotlivých hodnotitelů.

### 7.1 ACFS

Metoda ACFS, v češtině nazvaná „Dynamické testování kognitivních funkcí u dětí“, je českou adaptací metody ACFS (Application of Cognitive Functions Scale) autorky Carol Lidz. Tato metoda dynamické diagnostiky vznikla v souladu s CBDA (Curriculum-based dynamic assessment), neboli na kurikulu založené dynamické diagnostice, což je model zdůrazňující propojení výuky a diagnostiky. Metoda ACFS, která je určena pro předškolní děti, je tedy vystavěna tak, aby sledovala kognitivní procesy klíčové pro dosažení školní zralosti. ACFS využívá princip pre-test – intervence – post-test, který odděluje fáze testování od intervenční fáze. Intervence je zařazena v každém subtestu zvlášť a vychází z Feuersteinovy zkušenosti zprostředkovaného učení (MLE) (Lidz & Krejčová, 2014).

V subtestu **Třídění** má dítě za úkol stavět, rozřazovat a třídit kostky různých barev, tvarů a velikostí. Je hodnocena schopnost třídit do skupin podle tří různých charakteristik – barvy, tvaru a velikosti. Maximální skóre je 12.

**Sekvenční sluchová paměť** spočívá v převyprávění poslouchaného příběhu. Hodnotí se počet prvků a správné pořadí událostí v příběhu. Maximální skóre je 16.

Subtest **Krátkodobá vizuální paměť** obsahuje dvě sady fotografií s obrázky dvou druhů. Dítě si má obrázky zapamatovat. Hodnocen je počet zapamatovaných obrázků a strategie zapamatování. Maximální skóre je 14.

**Doplňování sekvenčních vzorců** spočívá v doplňování správného dílku do řady, která je sestavena v určitém pořadí. Je hodnocena správnost odpovědi, zda se jednalo o první, nebo druhý pokus a správné zdůvodnění odpovědi. Maximální skóre je 18.

**Přejímání perspektivy** je doplňkový subtest, v kterém má dítě převzít roli učitele a vysvětlit, jak nakreslit obrázek. Hodnotí se úroveň popisu a instrukcí na základě uvedených výroků. Maximální skóre je 16.

**Verbální plánování** je doplňkový subtest, ve kterém je úkolem dítěte popsat přípravu chleba či rohlíku, případně jiné aktivity. Je hodnocena propracovanost a struktura popisu a užívání slov týkajících se pořadí. Maximální skóre je 15 (Lidz & Krejčová, 2014).

**Intervenční fáze** probíhá vždy mezi pre-testem a post-testem každého subtestu. Odvíjí se od hodnotitelova pozorování výkonu a postupů testovaného v pre-testu a vychází ze zkušenosti zprostředkovaného učení. Jejím záměrem je nejen si společně s dítětem ukázat řešení úkolu, ale především ho přivést k aktivnímu používání pojmů, postupů a strategií řešení a podnítit tak strukturální kognitivní změny (Lidz & Krejčová, 2014).

V rámci intervenční fáze je většinou využíváno paralelních variant daných úkolů s odlišným podnětovým materiálem a výkon během intervence není hodnocen. Intervencí je navazováno na výkon dítěte v pre-testu. Důležitou složkou je osvojování postupů práce v jednotlivých subtestech a zprostředkování strategií řešení dítěti. Například v subtestu Třídění jde o kategorizaci, u Sekvenční sluchové paměti je k tréninku používána vizualizace informací, u Krátkodobé vizuální paměti napomáhá detailní prozkoumání podnětů a jejich třídění do skupin, u Doplnění sekvenčních vzorců jsou postupy intervence zaměřeny na uvědomění si sekvence. Detailnější postupy jsou popsány v manuálu metody a shrnuty do bodů v záznamovém archu, kde hodnotitel rovněž zaznamenává formu a intenzitu intervence pro každý subtest (Lidz & Krejčová, 2014).

**Fáze vyhodnocení** je popsána v manuálu metody pro každý subtest zvlášť a odvíjí se od povahy úkolu. Hodnotitel přiděluje skóre pro fázi pre-testu a post-testu u každého subtestu podle dosažené úrovně zvládnutí úkolu. Konkrétní vodítka pro přidělování skóre jsou obsažena v záznamovém archu.

## **7.2 Dotazník**

Metodu dotazníku ve formě tabulky jsme použily k získání údajů o administrátorech. Dotazovaly jsme se otevřenými otázkami na následující charakteristiky: dosažené vzdělání, obor, pracoviště, věk hodnotitele v době administrace metody a délka praxe v době administrace metody. Odpověď vystihující tyto charakteristiky dotázaní doplnili ke svému jménu do tabulky, která jim byla zaslána e-mailem.

## 8 Výzkumný vzorek

### 8.1 Administrátoři metody

Oslovily jsme všechny administrátory metody ACFS, kteří se podíleli na lokalizaci této metody v České republice. Z oslovených 16 administrátorů nám odpovědělo 15, kteří nám zaslali vyplněné údaje, na které jsme se dotázaly.

Jednalo se o 15 odborníků z oboru, nebo kombinace oborů, psychologie, speciální pedagogika, logopedie a učitelství pracujících v poradenských zařízeních, v soukromé klinické praxi, nestátní neziskové organizaci nebo škole. Vzorek těchto 15 dotázaných tvořily ženy ve věku 25–50 let s délkou praxe 0–33 roků v době provádění lokalizace. Všechny až na jednu hodnotitelku, která byla v posledním ročníku VŠ, měly ukončené vysokoškolské vzdělání. Všechny hodnotitelky byly v rámci lokalizace v této metodě vyškoleny.

Níže uvádíme zastoupení oborů, pracovišť a přehled věku a délky praxe administrátorů.

obory	četnost
psychologie	5
speciální pedagogika	2
učitelství	2
psychologie a pedagogika	4
psychologie a speciální pedagogika	1
Učitelství a speciální pedagogika	1

Tab. 1 Zastoupení oborů u skupiny administrátorů

typ pracoviště	četnost
poradenské zařízení	8
nezisková organizace	4
soukromá praxe	1
mateřská/základní/střední škola	2

Tab. 2 Zastoupení typů pracovišť

ID administrátora	věk	délka praxe
1	25	0
2	46	2
3	30	5
4	40	14
5	50	22
6	42	5
7	49	30
8	28	1,5
9	49	26
10	27	5
11	37	9
12	33	10
13	50	33
14	44	21
15	32	8

**Tab. 3 Věk a délka praxe administrátorů**

Údaje jsou uváděny v jednotkách let z doby provádění lokalizace.

## **8.2 Vzorek diagnostikovaných dětí**

Lokalizace byla prováděna na souboru 366 dětí ve věku 3-6 let. Z této skupiny 298 dětí nemělo stanovenou žádnou diagnózu, 37 dětí mělo stanovenou diagnózu z kategorie narušená komunikační schopnost a 30 dětí spadalo do skupiny jiné diagnózy, kam patřily diagnózy PAS, syndrom ADHD, nerovnoměrný kognitivní vývoj a navíc i bilingvní výchova (Lidz & Krejčová, 2014).

Pro účely této práce bylo do zpracování dat zahrnuto 341 dětí, které byly diagnostikovány 15 administrátory, kteří se zapojili do tohoto projektu.

## 9 Průběh výzkumu

Nejprve jsme oslovily odborníky, kteří jsou v České republice proškoleni v metodě ACFS, s žádostí o spolupráci na původním projektu této diplomové práce, jehož záměrem bylo prověřit reliabilitu této metody. Z 19 oslovených odpovědělo (i po opakovaném kontaktování) celkem 10 odborníků. Obsahem všech odpovědí bylo sdělení, že metodu ACFS v současné době nevyužívají a nemají naplánovaná žádná další vyšetření touto metodou. Zároveň pět z nich velmi ochotně nabídlo, že v případě, že by se situace změnila a naplánovali by vyšetření metodou AFCS, sami se ozvou a budou velmi rádi spolupracovat. Nikdo z nich se nicméně následně neozval. Z těchto důvodů byl projekt změněn a zadání práce upraveno do podoby současného znění. Okolnosti související s touto situací jsou komentovány v předmluvě práce. Dále se věnujeme průběhu nového diplomového projektu.

Pro tento výzkumný projekt byla použita data z české lokalizace metody ACFS provedené v roce 2014. Lokalizace se zúčastnilo 16 administrátorů, kteří diagnostikovali celkem 390 dětí, ze kterých bylo 366 zahrnuto do souboru zpracovaného v rámci lokalizace. Všichni tito administrátoři byli osloveni e-mailem s žádostí o další spolupráci. Byla jim zaslána tabulka, s prosbou o vyplnění údajů a zaslání zpět. Ti, kteří nereagovali na první žádost, byli osloveni opakovaně. Celkem zaslalo zpět vyplněné údaje a účastnilo se tak tohoto výzkumného projektu patnáct administrátorů z šestnácti, kteří dohromady diagnostikovali 341 dětí.

V datovém souboru z lokalizace ACFS bylo možné přesně identifikovat, který administrátor vyšetřil jednotlivé děti a podle toho byly údaje získané od administrátorů roztrženy. Údaje o administrátorech byly anonymizovány přiřazením identifikačního čísla každému z administrátorů. Následně byla data prozkoumána a statisticky analyzována.



## 10 Výsledky

V této kapitole budeme prezentovat výsledky analýzy dat, které jsme zpracovávaly v rámci tohoto diplomového projektu. Výsledky budou předkládány v pořadí, ve kterém byly formulovány výzkumné otázky.

Nejdříve se tedy budeme zabývat otázkou, zda existuje významný **vztah mezi hodnotiteli a skóry změny** u jednotlivých škál metody ACFS. Pro prozkoumání tohoto vztahu za kontrolování vlivu počáteční hodnoty (výsledku v pre-testu) byla zvolena metoda analýzy kovariance, kterou vzhledem k robustnosti této metody při velkém výběru a přibližně stejném rozsahu výběrů jednotlivých podskupin považujeme za vhodnou, i přes ne zcela normální rozložení závislé proměnné – v našem případě skóru změny u jednotlivých subtestů (viz Příloha 2).

Níže prezentujeme počet dětí testovaných danými hodnotiteli.

ID administrátora	N
1	25
2	24
3	15
4	13
5	26
6	25
7	25
8	25
9	25
10	25
11	25
12	12
13	26
14	25
15	25

Tab. 4 Rozsah výběrů pro jednotlivé hodnotitele

Zde uvádíme počet dětí, u kterých bylo v daných subtestech možné vypočítat skóre změny. U některých nebylo možné skóre změny spočítat, jelikož nebyl administrován post-test z důvodu dosažení velmi vysokého skóre v pre-testu, případně nebyly administrovány doplňkové subtesty z důvodu nízkého věku dítěte.

ID administrátora	N					
	Třídění	Sekvenční sluchová paměť	Krátkodobá vizuální paměť	Doplňování sekvenčních vzorců	Přejímání perspektivy	Verbální plánování
1	12	25	18	11	18	11
2	12	17	19	4	17	11
3	15	15	14	8	6	8
4	13	13	12	9	8	6
5	19	23	19	11	15	12
6	11	18	14	8	7	7
7	22	25	23	25	4	2
8	23	24	14	6	13	4
9	18	16	11	5	8	8
10	16	23	15	6	4	2
11	23	23	13	5	20	12
12	7	12	6	8	7	6
13	15	23	26	17	25	25
14	22	22	15	16	8	2
15	24	25	14	9	6	4

**Tab. 5 Rozsahy výběrů v daných subtestech**

Výsledky testování, prezentované v tabulce č. 6, ukázaly statisticky významný vztah mezi faktorem administrátora a závislé proměnné skóre změny u všech subtestů i po vyloučení vlivu počáteční hodnoty pre-testu, ačkoliv hodnoty pre-testu měly rovněž signifikantní význam pro velikost změny u všech subtestů. Výsledky ukazují signifikantní efekt v subtestu Třídění pro faktor administrátora  $F(14, 236) = 4,34$ ;  $p < 0,001$  a pro kovarianci pre-testu  $F(1, 236) = 100,58$ ,  $p < 0,001$ ; v subtestu Sekvenční sluchová paměť pro administrátora  $F(14, 288) = 4,48$ ,  $p < 0,001$  a pro pre-test  $F(1, 288) = 40,90$ ,  $p < 0,001$ ;

v subtestu Krátkodobá vizuální paměť pro administrátora  $F(14, 217) = 5,70$ ,  $p < 0,001$  a pro pre-test  $F(1, 217) = 53,20$ ,  $p < 0,001$ ; v subtestu Doplnování sekvenčních vzorců  $F(14, 132) = 3,37$ ,  $p < 0,001$  a pro pre-test  $F(1, 132) = 40,06$ ,  $p < 0,001$ ; v subtestu Přijímání perspektivy  $F(14, 150) = 3,44$ ,  $p < 0,001$  a pro pre-test  $F(1, 150) = 148,45$ ,  $p < 0,001$ ; v subtestu Verbální plánování  $F(14, 104) = 2,50$ ,  $p < 0,005$  a pro pre-test  $F(1, 104) = 22,47$ ,  $p < 0,001$ .

To znamená, že zamítáme nulovou hypotézu, že mezi průměry skóre změny dílčích subtestů není u probandů testovaných různými administrátory významný rozdíl. Přijímáme alternativní hypotézu, že průměrné skóre změny dílčích subtestů se významně liší u probandů testovaných různými administrátory.

závislá proměnná	faktor	F	Sig.
	kovariance		
skór změny v Třídění	administrátor	4,34	,000
	pre-test Třídění	100,58	,000
skór změny v Sekvenční sluchové paměti	administrátor	4,48	,000
	pre-test Sekvenční sluchové paměti	40,90	,000
skór změny v Krátkodobé vizuální paměti	administrátor	5,70	,000
	pre-test Krátkodobé vizuální paměti	53,20	,000
skór změny v Doplnování sekvenčních vzorců	administrátor	3,37	,000
	pre-test Doplnování sekvenčních vzorců	40,06	,000
skór změny v Přijímání perspektivy	administrátor	3,44	,000
	pre-test Přijímání perspektivy	148,45	,000
skór změny ve Verbálním plánování	administrátor	2,50	,004
	pre-test Verbální plánování	22,47	,000

**Tab. 6** Prověření vztahu pomocí analýzy kovariance

Testováno na hladině významnosti  $\alpha = 0,05$ .

Dále se věnujeme podrobnějšímu rozboru rozdílů mezi administrátory. Post-hoc analýza ukázala signifikantní rozdíly mezi konkrétními dvojicemi hodnotitelů v dílčích subtestech, jak znázorňuje tabulka č. 7.

subtest	dvojice hodnotitelů	rozdíl průměrů	Sig.
Třídění	1–7	2,70	,040
	2–7	3,45	,001
	5–7	3,48	,000
	8–7	2,58	,004
	9–7	3,05	,002
	11–7	3,37	,000
	15–7	3,13	,000
Sekvenční sluchová paměť	5–3	4,65	,017
	5–4	4,81	,019
	5–7	4,81	,001
	5–15	1,06	,039
	9–7	4,52	,015
	14–3	4,98	,007
	14–4	5,14	,008
	14–7	5,15	,000
Krátkodobá vizuální paměť	1–3	3,08	,001
	1–15	2,76	,006
	2–3	3,02	,001
	2–15	2,70	,008
	5–3	3,18	,000
	5–15	2,86	,002
	7–3	2,89	,001
	7–15	2,57	,009
	11–3	2,95	,006
	13–3	4,05	,000
	13–6	2,27	,041
Doplňování sekvenčních vzorců	5–4	5,43	,039
	5–14	5,75	,003
Přejímání perspektivy	2–13	1,82	,038
	2–15	2,23	,042
	7–3	3,85	,037
	7–11	3,36	,021

	7–13	3,31	,024
	7–15	4,21	,010
Verbální plánování	13–3	4,75	,012

**Tab. 7 Významné rozdíly mezi dvojicemi hodnotitelů pro dané subtesty**

Testováno na hladině významnosti  $\alpha = 0,05$ .

Výše uvedená tabulka podává přehled statisticky významných rozdílů ve skóru změny mezi hodnotiteli. V subtestu Třídění se u hodnotitele č. 7 ukázaly statisticky signifikantně výrazně nižší hodnoty změny, než u hodnotitelů 1, 2, 5, 8, 9, 11 a 15. V subtestu Sekvenční sluchová paměť byly výrazně vyšší hodnoty změny u hodnotitelů 5, 9, 14 a výrazně nižší u hodnotitelů 3, 4, 7, 15. V subtestu Krátkodobá vizuální paměť se objevily signifikantní rozdíly mezi několika hodnotiteli. Výrazně vyšší hodnoty změny byly u hodnotitelů 1, 2, 5, 7, 11 a 13 a výrazně nižší u hodnotitelů 3, 6, 15, přičemž u hodnotitele 3 byla průměrná hodnota dokonce záporná. Ve čtvrtém subtestu Doplnění sekvenčních vzorců byly statisticky významné rozdíly mezi hodnotiteli 5 a 4, a 5 a 14, kde se u hodnotitele č. 5 objevily vyšší hodnoty a u hodnotitelů 4 a 14 nižší hodnoty. V doplňkovém subtestu Přejímání perspektivy byly významně vyšší hodnoty změny u hodnotitele 2 oproti hodnotitelům 13 a 15, a také u hodnotitele 7 oproti hodnotitelům 3, 11, 13 a 15 (u hodnotitele 15 byla průměrná hodnota změny záporná). V posledním subtestu Verbální plánování se statisticky významný rozdíl objevil pouze u dvojice hodnotitelů 13 a 3, kde hodnotitel č. 13 měl významně vyšší hodnoty změny než hodnotitel 3, u kterého byla v tomto subtestu opět průměrná hodnota skóru změny záporná.

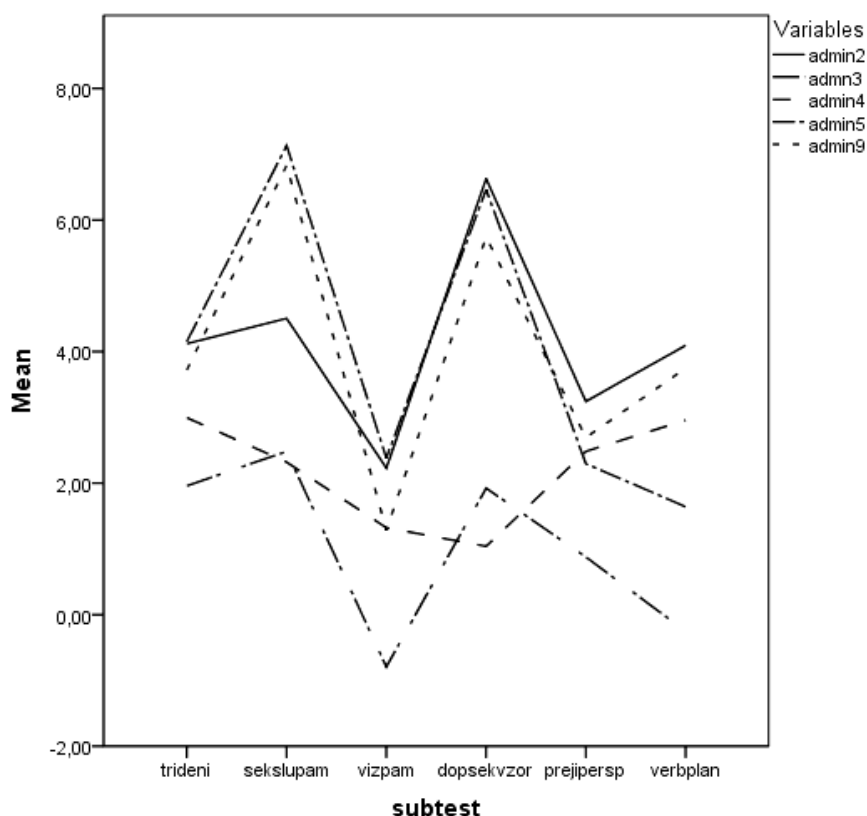
Další otázkou je, zda se tyto rozdíly mezi hodnotiteli shodují nebo podobají mezi subtesty, tedy zda se u hodnotitelů objevují **rozdílné tendence napříč jednotlivými škálami** metody ACFS. Z výše uvedeného výčtu je patrné, že se rozdíly mezi hodnotiteli napříč jednotlivými škálami liší. Z důvodu nízkého počtu hodnotitelů ( $N=15$ ) je možné tuto otázku prozkoumat pouze kvalitativně. Pro přehled však uvádíme v tabulce 8 četnost výše zmíněných rozdílů u jednotlivých hodnotitelů.

ID administrátora	četnost významně vyšších hodnot	četnost významně nižších hodnot
1	2	0
2	3	0
3	1	4
4	0	2
5	3	0
6	0	1
7	2	2
8	1	0
9	2	0
10	0	0
11	2	1
12	0	0
13	2	1
14	1	1
15	1	3

**Tab. 8 Četnosti výrazných rozdílů u hodnotitelů**

Z kvalitativního rozboru můžeme určit, u kterých hodnotitelů se vícekrát objevovaly vyšší skóry změny napříč subtesty a u kterých nižší. Vzhledem tomu, že u žádného administrátora nebyly statisticky významně vyšší nebo nižší skóry změny ve srovnání s ostatními u většiny subtestů, s výjimkou administrátora č. 3, u kterého byly čtyři významně nižší průměry a jeden významně vyšší, nemůžeme usuzovat na významný rozdíl mezi konkrétními administrátory napříč různými subtesty.

Přesto můžeme pozorovat mezi těmito hodnotiteli mírně rozdílné tendence. Následující graf (Obr. 1) zobrazuje hodnocení administrátorů, u kterých se ve srovnání s ostatními hodnotiteli vícekrát objevil statisticky významně vyšší nebo nižší rozdíl ve skóru změny hodnocených probandů. Z grafu je patrné, že u těchto hodnotitelů můžeme pozorovat napříč subtesty tendence vyšších skórů změny u hodnotitelů 2, 5 a 9 a tendence nižších skórů změny u hodnotitelů 3 a 4. Z dostupných dat však vyplývá, že tyto tendence nejsou statisticky významné. Možné příčiny a souvislosti budeme komentovat v diskusi.



Obr. 1 Graf průměrných skóre změny v daných subtestech pro administrátory 2, 3, 4, 5 a 9

Na závěr prozkoumáme, které **charakteristiky** se objevují u hodnotitelů, u kterých jsme pozorovaly opakované rozdíly v hodnocení ve srovnání s ostatními administrátory ve dvou nebo více subtestech. Z kvalitativního posouzení charakteristik těchto administrátorů nám nejsou zřetelné žádné nápadnosti mezi hodnotiteli s opačnými tendencemi. Nepozorovaly jsme mezi těmito skupinami hodnotitelů žádné výrazné rozdíly. V obou skupinách jsou hodnotitelé s různými obory a převažující typ pracoviště je napříč skupinami poradenské zařízení. Malé rozdíly můžeme pozorovat v kategorii věk a délka praxe, kde pro skupiny hodnotitelů s průměrnými vyššími skóre změny je věkové rozmezí spíše odpovídající pozdní dospělosti a délka praxe ve dvou případech ze tří přesahuje 20 let, zatímco u hodnotitelů s průměrnými nižšími skóre změny odpovídá věk střední dospělosti a délka praxe je menší než u předchozí skupiny.

S ohledem na povahu dat nebylo možné vliv konkrétních charakteristik hodnotitelů na velikost změny prozkoumat jiným způsobem, což více komentujeme v diskusi.

	ID administrátora	obor	pracoviště	věk	délka praxe
vyšší skóry změny	2	psychologie	poradenské zařízení	46	2
	5	učitelství se zaměřením speciální pedagogika	poradenské zařízení	50	22
	9	speciální pedagogika	poradenské zařízení	49	26
nižší skóry změny	3	psychologie	soukromá klinická praxe	30	5
	4	psychologie, učitelství	poradenské zařízení	40	14

**Tab. 9** Charakteristiky hodnotitelů s výraznějšími tendencemi



## 11 Diskuse

V diskusi nejprve stručně shrneme výzkumný projekt a cíl této diplomové práce a následně se budeme věnovat výsledkům získaným na základě zpracování dat. Výsledky se pokusíme zasadit do kontextu dosavadních poznatků a komentovat souvislosti, případně konfrontovat s dostupnými studiemi. Popíšeme limity výzkumného projektu a případná doporučení pro další výzkum. Na základě výstupů z této studie i poznatků podaných v teoretické části práce se pokusíme diskutovat ideální model prověřování spolehlivosti a stability dynamické diagnostiky.

**Cílem** této práce bylo jednak zmapovat způsoby a možnosti prověřování reliability a validity metod dynamické diagnostiky, čemuž jsme se věnovaly v teoretické části práce, prozkoumat vztah hodnotitelů a výsledků dynamické diagnostiky v empirické části práce, a na závěr navrhnout optimální přístup prověřování reliability a validity u metod tohoto typu.

Výzkumný projekt navazoval na lokalizaci metody ACFS v českém prostředí a jeho záměrem bylo prozkoumat vztah charakteristik hodnotitelů, kteří se na této lokalizaci podíleli, a velikosti dosažené změny testovaných v jednotlivých subtestech. Pomocí analýzy kovariance jsme testovaly závislost výsledné změny na faktoru hodnotitele za současné kontroly počáteční hodnoty pre-testu.

Na základě **výsledků** výše zmíněného testování jsme zamítly nulovou hypotézu, že mezi průměry skóre změny dílčích subtestů není u probandů testovaných různými administrátory významný rozdíl a přijaly jsme alternativní hypotézu, že průměrné skóre změny se u všech dílčích subtestů významně liší u probandů testovaných různými administrátory. Výsledky následné analýzy však ukázaly, že v každém ze subtestů se navzájem výrazně lišily jiné dvojice hodnotitelů. Přesto, že u některých hodnotitelů se signifikantně vyšší nebo nižší skóre změny opakovaly ve více subtestech, výsledky nenaznačují, že by některý z hodnotitelů skóroval výrazně odlišně ve všech nebo ve většině subtestů. Nepředpokládáme tedy významně odlišné tendence ve skórování mezi hodnotiteli v rámci celé metody.

Hodnotitele, u kterých se významné odlišnosti objevily ve více subtestech, jsme se nicméně pokusily dále kvalitativně prozkoumat. Graf jejich průměrného skórování ukázal u některých hodnotitelů vyšší a u některých nižší průměrné skóre napříč metodou. Jak bylo řečeno, tyto rozdíly však nebyly signifikantní, proto je nazýváme mírnými tendencemi.

Tyto tendence není možné zcela vysvětlit pomocí charakteristik hodnotitelů. Na základě kvalitativního rozboru, který ukázal určité, ač malé rozdíly v kategoriích věk a délka praxe (kde pro skupiny hodnotitelů s průměrnými vyššími skóry změny věkové rozmezí spíše odpovídalo pozdní dospělosti a délka praxe ve dvou případech ze tří přesahovala 20 let, zatímco u hodnotitelů s průměrnými nižšími skóry změny věk odpovídal střední dospělosti a délka praxe byla nižší), se můžeme domnívat, že tyto charakteristiky by mohly s administrováním a skórováním metody souviset. Tuto otázku by bylo nutné podrobněji prozkoumat v další studii.

Zmíněné tendence by však také mohly souviset s charakteristikami diagnostikovaných dětí. Souvislost odlišných výstupů dynamické diagnostiky s charakteristikami testovaných dětí by odpovídala i výsledkům dosavadních výzkumů. Například byla prokázána větší změna mezi pre-testem a post-testem u dětí s nižší inteligencí než u dětí s vyšší inteligencí (Stevenson et al., 2013). Také Tzuriel 1989 (podle Tzuriel 2000) ve své studii prokázal větší zlepšení od pre-testu k post-testu u znevýhodněných dětí než u dětí nadaných. Míra zlepšení znevýhodněných dětí oproti nadaným byla také větší u složitějších a komplexnějších úkolů. To potvrzuje, že zprostředkování je efektivnější, a tudíž i potřebnější u komplexních a abstraktních úkolů, což by mohlo souviset i s rozdíly v dosaženém hodnocení změny mezi jednotlivými subtesty, které ukázaly výsledky naší studie.

Výsledky našeho testování jsou ve shodě se studií zkoumající závislost výstupů LPAD na různorodosti administrátorů a jejich chování v průběhu administrace. Tato studie ukázala rozdíly v administrování metody mezi hodnotiteli, nicméně nebylo možné prokázat, zda zmíněné rozdíly významně ovlivňují výstupy testování (Burns, 1996, podle Sternberg & Grigorenko, 2002). Tato otázka tak zůstává otevřená a bylo by vhodné se jí věnovat v dalších studiích.

Výstup našeho výzkumu, který zřetelně neukázal na význam konkrétních charakteristik hodnotitele pro výsledné hodnocení dynamické diagnostiky, je také v částečném souladu se studií autorů Hansen et al. (2009). Zmíněná studie prověřovala charakteristiky hodnotitele v souvislosti s reliabilitou metody na základě shody posuzovatelů a následné analýzy možných souvislostí s charakteristikami hodnotitelů. Autoři analyzovali charakteristiky, jako jsou uplynulá doba od ukončení studia, nebo zkušenosti klinické praxe. Analýza neprokázala, že by tyto charakteristiky ovlivňovaly

výsledky shody hodnotitelů, což je v souladu s výstupem našeho rozboru charakteristik u jednotlivých hodnotitelů.

Na druhou stranu zmíněná možnost souvislosti skórování s charakteristikami věk a délka praxe, která by však musela být více prozkoumána, by odpovídala názorům zdůrazňujícím důležitost zkušeností, zaškolení a tréninku v používání metody. Důležitost intenzivního tréninku v metodě zdůrazňují a doporučují vzhledem k náročnosti metody a nárokům kladeným na dovednosti a odbornost hodnotitele, jak bylo podrobněji popsáno v teoretické části práce, například autoři Hansen et al. (2009), Grigorenko (2009), Tzuriel a Haywood (2002) a Lidz & Krejčová (2014).

Další dostupné studie prověřující vztah výstupů dynamické diagnostiky a hodnotitelů, případně jejich charakteristik nám nejsou známy.

Nyní se budeme věnovat **limitům** této práce. Jedním z limitů našeho výzkumu je malá variabilita charakteristik hodnotitelů, které jsme zkoumaly. Ta souvisí s počtem hodnotitelů (N=15), kteří se lokalizace účastnili. S ohledem na rozložení zkoumaných charakteristik ve spojení s daty z lokalizace metody ACFS nebylo možné vliv jednotlivých charakteristik hodnotitelů na velikost změny prozkoumat zvlášť. Z toho důvodu nemůžeme dostatečně posoudit význam konkrétních charakteristik v souvislosti vztahu hodnotitele a výstupu testování.

Jiným limitem může být rozložení zkoumaných dat z lokalizace, které nebylo pro naše účely optimální, nicméně by mělo být vzhledem k robustnosti použité metody analýzy dostačující. Přesto je třeba na výsledky nahlížet s vědomím, že rozložení a rozsah zkoumaného výběru mohou být pro výstupy analýzy intervenujícími faktory.

Vliv na výsledné skóry změny, a tudíž i na výsledky naší studie, může mít i „stropový efekt“, kdy je hodnota dosaženého skóru limitována maximálním možným počtem bodů. To se mohlo projevit jednak v hodnotě skóru změny, kdy děti s vyššími hodnotami v pre-testu neměly takovou možnost zlepšení jako děti s nižšími skóry, a také se to projevilo omezením počtu probandů, u kterých bylo možné vypočítat skór změny vzhledem k dosažení velmi vysokých hodnot u některých dětí již v pre-testu.

Hlavním limitem této práce je skutečnost, že vzhledem k okolnostem komentovaným v předmluvě, metoda ACFS není v současnosti v České republice v praxi dostatečně využívána, tudíž nemohl být realizován původní výzkumný projekt s cílem prověřit reliabilitu metody ACFS.

Níže se pokusíme s ohledem na výše popsané výsledky a zkušenosti zprostředkované studii a popsané v teoretické části podat **návrh optimálního přístupu** k prověřování reliability a validity metody dynamické diagnostiky. Nejdříve se budeme věnovat způsobu prověřování reliability a poté prověřování validity metod dynamické diagnostiky.

V první řadě je třeba brát v potaz přístup k dynamické diagnostice konkrétní metody, kterou chceme prověřovat. Jak bylo řečeno v teoretické části, v pohledu na prověřování psychometrických charakteristik se od sebe výrazně liší klinicky orientovaná a výzkumně orientovaná dynamická diagnostika. S tím souvisí i odlišné principy testování – „pre-test – intervence – post-test“ a průběžné poskytování intervence v průběhu celého vyšetření. U metod, kde není fáze intervence jasně ohraničena, je velmi obtížné reliabilitu prověřovat. Feuerstein, Rand, Jensen, Kaniel a Tzuriel (1987, podle Tzuriel, 2001), dokonce uvádějí, že měření reliability je u dynamické diagnostiky nevhodné, jelikož hlavním cílem dynamické diagnostiky je podnítit změnu a tudíž by neměla být charakterizována neměnností a stabilitou výsledků reprezentujících reliabilitu. Nutno podotknout, že autoři zde argumentují proti tak zvané mezipoložkové reliabilitě týkající se testovaného subjektu či test – re-testové reliabilitě, zatímco tyto argumenty se netýkají reliability shody mezi hodnotiteli diagnostikujícími stejného jedince, která by měla být i u dynamické diagnostiky vysoká (Tzuriel, 2001).

Z toho důvodu je u prověřování metod dynamické diagnostiky také důležité přemýšlet o druhu reliability. Jak bylo zmíněno, reliabilita dynamické diagnostiky je problematickou otázkou již z principu podněcování záměrné změny charakteristik, které jsou hodnoceny. Částečným řešením je proto, například také podle autorů Tzuriel a Haywood (2002), potvrzení vysoké reliability nástrojů použitých v procesu dynamické diagnostiky ve statickém modu, jak tomu bylo například u některých metod popsaných v teoretické části práce. Částečně má také velký význam metoda odhadu reliability pomocí shody hodnotitelů, která je k těmto účelům často využívána.

Výsledky naší studie, které naznačují, že mezi hodnotiteli existují rozdíly ve skórování a hodnocení změny, nicméně nevysvětlují, které charakteristiky hodnotitelů jsou pro tyto rozdíly významné, svědčí pro důležitost měření shody hodnotitelů u metod dynamické diagnostiky. To je v souladu s tvrzením autorů Beyer a Magnusson (2003, podle Hansen et al., 2009), kteří uvádějí, že vysoká míra subjektivity implikuje sníženou

reliabilitu metody, a proto má shoda mezi hodnotiteli pro prověřování charakteristik dynamické diagnostiky velice důležitý význam.

Tento způsob prověření považujeme za optimální i pro námi zkoumanou metodu ACFS, u které není z důvodů jednoúkolových subtestů možné stanovit vnitřní konzistenci testu a vzhledem k výše popsaným specifikům této metody je tedy posouzení reliability na základě shody hodnotitelů jediným možným a současně žádoucím způsobem. Potvrdila se tak důležitost záměru původního projektu, který měl za cíl reliabilitu metody ACFS tímto způsobem prověřit.

U prověřování validity metod dynamické diagnostiky je, s odkazem na teoretickou část práce, podstatné brát vždy v potaz cíl a záměr testování. Proto považujeme za důležité zdůraznit, že metoda by měla být validizována vždy pro konkrétní populaci, pro kterou má být používána. Vzhledem k hlavnímu záměru dynamické diagnostiky podnítit pozitivní změnu u testovaného, je vhodné hodnotit, zda metoda skrze zprostředkované učení tuto změnu skutečně podněcuje. Pro takové hodnocení je optimální model prověření, který srovnává experimentální skupinu, která podstoupila intervenci, a kontrolní skupinu, která intervenci nepodstoupila. Tento model použila také Drápalová (2014) pro prověření metody ACFS ve studii, kterou považujeme za velmi kvalitní a která je také více popsána v teoretické části práce.

Pokud je v kontextu hodnocení potenciálu cílem metody předvídat dosažení určitých výsledků či školního úspěchu jedince, má zde svůj význam i prediktivní validita. Zde je však velmi důležité dbát na zvolení vhodného srovnávacího kritéria, které by mělo odpovídat cílům validizace a na základě kterého by měly být výsledky validizace také interpretovány.

Jako vhodný způsob prověření validity dynamické diagnostiky se jeví i postup, který navrhuji Tzuriel a Haywood (2002), který spočívá v použití statické i dynamické diagnostiky pro jednu skupinu dětí, a pouze statického normativního testu pro druhou skupinu. Část dětí ve skupině s dynamickou diagnostikou by podstoupila včleněnou intervenci a část ne. Očekávání by byla taková, že statické testy by byly úspěšným prediktorem školního úspěchu u skupiny dětí, které podstoupily pouze statickou diagnostiku a u podskupiny dynamické diagnostiky bez začleněné intervence. U skupiny dětí s dynamickou diagnostikou s včleněnou intervencí by měly být statické testy signifikantně horším prediktorem úspěchu (Tzuriel & Haywood, 2002).

Zajímavou myšlenkou, v kontextu prediktivní validity školního úspěchu, je začlenění do dynamického testování prototypických úloh odpovídajícím školním aktivitám (Sternberg & Grigorenko, 2002). Autoři uvádějí, že rozdílné styly učení mohou korespondovat s rozdílnými typy úkolů a dynamické, na studenta orientované úlohy často korespondují s více otevřeným a progresivním stylem, než je tomu u tradičních úloh častěji používaných ve škole. Proto by pro zjišťování prediktivní validity měly být použity úkoly srovnatelného typu (Sternberg & Grigorenko, 2002).

Další způsoby prověřování validity (viz kapitola Zdroje důkazů o validitě) považujeme rovněž za důležité a mohou být z našeho pohledu zdrojem cenných informací. Za ideální přístup proto považujeme kombinaci různých zdrojů důkazů o validitě, které mohou dohromady podat komplexnější pohled na validitu metody. Zde jsme zmínily pouze ty, které u metod dynamické diagnostiky považujeme za důležité zdůraznit, ostatní byly popsány v teoretické části práce.

## **Závěr**

Reliabilita a validita metod dynamické diagnostiky je stále aktuálním a nedořešeným tématem, které poskytuje mnoho prostoru k dalším výzkumným studiím. Ačkoliv se jednotlivé názory a přístupy k dynamické diagnostice a prověřování jejích psychometrických charakteristik odlišují, přispívají k podnětné diskusi zdůrazňující důležitost tohoto tématu.

V naší práci jsme se pokusily vystihnout klíčová témata této problematiky a zprostředkovat poznatky z dosavadních studií, jež se daným tématem zabývají. V empirické části jsme se blíže zaměřily na hodnotitele a pokusily jsme se přispět k prozkoumání jeho významu v dynamické diagnostice.

Výsledky analýzy dat poukazují na skutečnost, že hodnotitel je významným faktorem pro velikost dosaženého skóru změny ve všech subtestech metody ACFS, což podporuje důležitost role hodnotitele v dynamické diagnostice. Nicméně jsme pozorovaly pouze mírné tendence ve skórování u několika hodnotitelů napříč subtesty, a z výsledků studie není možné jednoznačně určit význam konkrétních charakteristik hodnotitelů. Tyto výsledky přispívají k dosavadním poznatkům o procesu dynamického testování a současně přinášejí otázky, které mohou být podnětné pro další studie, jak je více komentováno v diskusi.

V neposlední řadě v diskusi navrhuje, na základě poznatků z dosavadních studií a výstupů naší práce, optimální přístup k prověřování reliability a validity metod dynamické diagnostiky. Zde zdůrazňujeme především význam odhadu reliability na základě shody posuzovatelů a důkazy validity založené na srovnání experimentální a kontrolní skupiny v kontextu záměru testování.

## Seznam použité literatury:

- Aranov, Z. (1999). *Validity and Reliability of the ACFS Behavior Observation Rating Scale*. Touro College.
- Babad, E. Y., & Budoff, M. (1974). Sensitivity and validity of learning-potential measurement in three levels of ability. *Journal of Educational Psychology*, 66(3), 439–447.
- Baek, S.-G., & Kim, K. J. (2003). The Effect of Dynamic Assessment Based Instruction on Children's Learning. *Asia Pacific Education Review*, 4(2), 189–198.
- Bensoussan, Y. (2002). *The Effectiveness of Mediation on Three Subtests of the Application of Cognitive Functions Scale, A Dynamic Assessment Procedure for Young Children*. Touro College.
- Binger, C., Kent-Walsh, J., & King, M. (2017). Dynamic Assessment for 3- and 4-Year-Old Children Who Use Augmentative and Alternative Communication: Evaluating Expressive Syntax. *Journal of Speech Language and Hearing Research*, 60(7), 1946–1958.
- Brandão, S., Cadime, I., Ribeiro, I., & Viana, F. L. (2014). Evaluating listening comprehension in children using a dynamic assessment approach: An exploratory study. *European Journal of Developmental Psychology*, 11(6), 746–753.
- Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The Predictive Validity of Dynamic Assessment: A Review. *The Journal of Special Education*, 41(4), 254–270.
- Calero, M. D., Belen, G. M. M., & Robles, M. A. (2011). Learning Potential in high IQ children: The contribution of dynamic assessment to the identification of gifted children. *Learning and Individual Differences*, 21(2), 176–181.
- Camilleri, B., & Botting, N. (2013). Beyond static assessment of children's receptive vocabulary: the dynamic assessment of word learning (DAWL). *International Journal of Language & Communication Disorders*, 48(5), 565–581.
- Carver, R. P. (1974). Two dimensions of tests: Psychometric and edumetric. *American Psychologist*, 29(7), 512–518.
- De Beer, M. (2010). A Modern Assessment Psychometric Approach to Dynamic Assessment. *Journal of Psychology in Africa*, 20(2), 241–246.
- Drápalová, A. (2014). *Možnosti dynamické diagnostiky u dětí v mateřských školách – Škála využití kognitivních funkcí (ACFS) Dynamic*. Univerzita Karlova, Praha.
- Dykeman, F. B. (2006). Alternative Strategies in Assessing Special Education Needs. *Education*, 127(2), 265–274.
- Elliott, J. (2003). Dynamic assessment in educational settings: realising potential. *Educational Review*, 55(1), 15–32.
- Ferjenčík, J. (2010). *Úvod do metodologie psychologického výzkumu*. Praha: Portál.
- Grigorenko, E. L. (2009). Dynamic Assessment and Response to Intervention: Two Sides of One Coin. *Journal of Learning Disabilities*, 42(2), 111–132.



- Hansen, T., Steultjens, E., & Satink, T. (2009). Validation of a Danish translation of an occupational therapy guideline for interventions in apraxia: A pilot study. *Scandinavian Journal of Occupational Therapy*, 16(4), 205–215.
- Hasson, N., Dodd, B., & Botting, N. (2012). Dynamic Assessment of Sentence Structure (DASS): design and evaluation of a novel procedure for the assessment of syntax in children with language impairments. *International Journal of Language & Communication Disorders*, 47(3), 285–299.
- Haywood, H. C., & Wingenfeld, S. A. (1992). Interactive Assessment As a Research Tool. *The Journal of Special Education*, 26(3), 253–268.
- Kelley, E. S. (2017). Measuring Explicit Word Learning of Preschool Children: A Development Study. *American Journal of Speech-Language Pathology*, 26(3), 961–971.
- Kester, E. S., Peña, E. D., & Gillam, R. B. (2001). Outcomes of Dynamic Assessment with Culturally and Linguistically Diverse Students: A Comparison of Three Teaching Methods Within a Test-Teach-Retest Framework. *Journal of Cognitive Education and Psychology*, 2(1), 37–54.
- Kozulin, A., & Rand, Y. (2000). *Experience of Mediated Learning: An Impact of Feuerstein's Theory in Education and Psychology*. New York: Pergamon.
- Lidz, C. S. (1991). *Practitioner's guide to dynamic assessment*. New York: The Guilford Press.
- Lidz, C. S. (2003). *Early Childhood Assessment*. New Jersey: John Wiley & Sons.
- Lidz, C. S. (2014). Leaning Toward a Consensus About Dynamic Assessment: Can We? Do We Want To? *Journal of Cognitive Education and Psychology*, 13(3), 292–307.
- Lidz, C. S., & Krejčová, L. (2014). *Dynamické testování kognitivních funkcí u dětí: ACFS-cz - Manuál*. Otrokovice: Propsyco.
- Lidz, C. S., & Peña, E. D. (1996). Dynamic Assessment: The Model, its Relevance as a Nonbiased Approach, and its Application to Latino American Preschool Children. *Language, Speech, and Hearing Services in Schools*, 27, 367–372.
- Luwel, K., Foustana, A., Papadatos, Y., & Verschaffel, L. (2010). The role of intelligence and feedback in children's strategy competence. *Journal of Experimental Child Psychology*, 108(1), 61–76.
- Malowitzky, M. (2001). *Investigation of the Effectiveness of the Mediation Portion of Two Subtests of the Application of Cognitive Functions Scale, a Dynamic Assessment Procedure for Young Children*. Touro College.
- Mata, S., & Calero, M. D. (2014). Dynamic assessment: the Spanish version of the Application of Cognitive Functions Scale. *The Spanish Journal of Psychology*, 17, 1–9.
- Missiuna, C., & Samuels, M. T. (1989). Dynamic Assessment of Preschool Children with Special Needs: Comparison of Mediation and Instruction. *Remedial and Special Education*, 10(2), 53–60.
- Naeini, J. (2014). A Comparative Study of the Effects of Two Approaches of Dynamic Assessment on the Reading Comprehension of Iranian EFL Learners. *International Journal of Applied Linguistics & English Literature*, 4(2), 54–66.

- Neubauer, A. C., Grabner, R. H., Freudenthaler, H. H., Beckmann, J. F., & Guthke, J. (2004). Intelligence and individual differences in becoming neurally efficient. *Acta Psychologica*, 116(1), 55–74.
- Peña, E., Iglesias, A., & Lidz, C. S. (2001). Reducing Test Bias Through Dynamic Assessment of Children's Word Learning Ability. *American Journal of Speech-Language Pathology*, 10(2), 138–154.
- Shurin, R. (1999). *Concurrent and Discriminant Validity of a Dynamic Assessment Procedure with Special Needs and Typical Preschool Children*. Touro College.
- Standardy pro pedagogické a psychologické testování. (2001). Praha: Testcentrum.
- Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: the nature and measurement of learning potential*. New York: Cambridge University Press.
- Stevenson, C. E., Hickendorff, M., Resing, W. C. M., Heiser, W. J., & de Boeck, P. A. L. (2013). Explanatory item response modeling of children's change on a dynamic test of analogical reasoning. *Intelligence*, 41(3), 157–168.
- Swanson, H. L. (1995). Using the cognitive processing test to assess ability: Development of a dynamic assessment measure. *School Psychology Review*, 24(4), 682–693.
- Tzuriel, D. (1990). *Children's Inferential Thinking Modifiability Test: Instruction Manual*. Ramat-Gan: School of Education, Bar-Ilan University.
- Tzuriel, D. (2000). Dynamic assessment of young children: Educational and intervention perspectives. *Educational Psychology*, 12(4), 385–435.
- Tzuriel, D. (2001). *Dynamic assessment of young children*. New York: Springer Science.
- Tzuriel, D., & Galinka, E. (2002). *Children's Conceptual and Perceptual Analogical Modifiability Test: Instruction Manual*. Ramat-Gan: School of Education, Bar-Ilan University.
- Tzuriel, D., & George, T. (2009). Improvement of analogical reasoning and academic achievements by the Analogical Reasoning Programme (ARP). *Educational and Psychology Child*, 26(3), 71–94.
- Tzuriel, D., & Haywood, H. C. (2002). Applications and Challenges in Dynamic Assessment. *Peabody Journal of Education*, 77(2), 40–63.
- Tzuriel, D., & Klein, P. S. (1985). The Assessment of Analogical Thinking Modifiability Among Regular, Special Education, Disadvantaged, and Mentally Retarded Children. *Journal of Abnormal Child Psychology*, 13(4), 539–552.
- Tzuriel, D., & Klein, P. S. (1990). *Children's Analogical Thinking Modifiability Test: Instruction Manual*. Ramat-Gan: School of Education, Bar-Ilan University.
- Urbánek, T., Denglerová, D., & Širůček, J. (2011). *Psychometrika: Měření v psychologii*. Praha: Portál.
- Van Der Aalsvoort, G. M., & Lidz, C. S. (2007). A Cross-Cultural Validation Study of the Application of Cognitive Functions Scale. *Journal of Applied School Psychology*, 24(1), 91–108.
- Whyte, J., & Hart, T. (2003). It's More Than a Black Box; It's a Russian Doll. *American Journal of Physical Medicine & Rehabilitation*, 82(8), 639–652.
- Wiedl, K. H., Mata, S., Waldorf, M., & Calero, M. D. (2014). Dynamic testing with native

and migrant preschool children in Germany and Spain, using the Application of Cognitive Functions Scale. *Learning and Individual Differences*, 35, 34–40.

## **Seznam příloh:**

Příloha 1: Žádost o spolupráci na původním projektu	I
Příloha 2: Normalita rozložení výběru z hlediska skóru změny subtestů	II